DOCUMENT RESUME

ED 241 793 CE 038 651

AUTHOR Baum, David R.; And Others
TITLE Specification of Training Simulator Fidelity: A Research Plan. Technical Report 558.
INSTITUTION Honeywell, Inc., Minneapolis, Minn. Systems and Research Center.
SPONS AGENCY Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
PUB DATE Feb 82
CONTRACT MDA-903-81-C-0214
NOTE 96p.; Product of the Simulation Systems Technical Area.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adult Education; *Job Training; Military Personnel; *Military Training; *Research Design; *Simulation
IDENTIFIERS Army; *Simulation Fidelity; *Training Effectiveness

ABSTRACT

This report presents a research plan to guide the determination of the empirical relationship between level of maintenance training simulator fidelity and training effectiveness. Chapter I describes data collection and analysis activities undertaken to provide guidance for fidelity decision making by the training simulator development community. These include a review of the empirical literature on training simulator fidelity; site visits and interviews at selected Army, Navy, and Air Force agencies; and a fidelity research issues workshop. The chapter also discusses issues related to the definitions of fidelity and fidelity specification and examines additional factors that determine device-training effectiveness. Chapter II presents the rationale for a framework within which a specific plan of fidelity research can be developed and conducted. The framework is developed through considerations of (1) operational definitions of fidelity, (2) economic sequencing of studies, (3) theoretical and empirical aspects of maintenance tasks and required skills, and (4) desired subject population characteristics. Chapter III presents potential research paradigms for studying the relationship between training simulator fidelity and transfer of training for perceptual-motor and cognitive tasks. General design considerations for the pilot studies are discussed. Appendixes include a description and analysis of the Army's training simulator development process and workshop details. (YLB)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED241793

Technical Report 558

# SPECIFICATION OF TRAINING SIMULATOR FIDELITY: A RESEARCH PLAN

David R. Baum, Deborah A. Smith and Stephen F. Hirshfeld
Honeywell Systems and Research Center

Gary A. Klein
Klein Associates

Robert W. Swezey
Sciences Applications, Inc

Robert T. Hays
Army Research Institute

SIMULATION SYSTEMS TECHNICAL AREA

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

U. S. Army
Research Institute for the Behavioral and Social Sciences

February 1982

Approved for public release; distribution unlimited.

CE031651

# U. S. ARMY RESEARCH INSTITUTE
# FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract
for the Department of the Army

Honeywell Systems and Research Center

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-TST, 5001 Eisenhower Avenue, Alexandria, Virginia 22333.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>Technical Report 558 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>SPECIFICATION OF TRAINING SIMULATOR FIDELITY: A RESEARCH PLAN | | 5. TYPE OF REPORT & PERIOD COVERED<br>Interim Report/March-October 1981<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>David R. Baum, Deborah A. Smith and Stephen F. Hirshfeld, Gary A. Klein, Robert W. Swezey, and Robert T. Hays | | 8. CONTRACT OR GRANT NUMBER(s)<br>MDA 903-81-C-0214 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Honeywell Systems and Research Center<br>2600 Ridgway Parkway<br>Minneapolis, Minnesota 55413 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>2Q162717A790 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>US Army Research Institute for the Behavioral and Social Sciences (PERI-IE), 5001 Eisenhower Avenue, Alexandria, VA 22333 | | 12. REPORT DATE<br>February 1982<br>13. NUMBER OF PAGES<br>84 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report)<br>Approved for public release; distribution unlimited | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number)<br>Simulation training<br>Simulators<br>Simulator fidelity<br>Fidelity guidance<br>Fidelity specification<br>Fidelity requirements<br>Fidelity research<br>Training device effectiveness<br>Maintenance training | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number)<br>This report presents a conceptual framework for identifying factors that may impact training simulator effectiveness. A research strategy is proposed for the empirical determination of necessary levels of training simulator fidelity. Preliminary studies consistent with this strategy are described to explore the effects of device fidelity on the transfer of training of perceptual-motor and cognitive maintenance tasks. | | |

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Technical Report 558

# SPECIFICATION OF TRAINING SIMULATOR FIDELITY: A RESEARCH PLAN

David R. Baum, Deborah A. Smith and Stephen F. Hirshfeld
Honeywell Systems and Research Center

Gary A. Klein
Klein Associates

Robert W. Swezey
Sciences Applications, Inc

Robert T. Hays
Army Research Institute

Submitted by:
Frank J. Harris, Chief
**SIMULATION SYSTEMS TECHNICAL AREA**

Approved by:
Milton S. Katz, Acting Director
TRAINING RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

February 1982

Army Project Number
2Q162717A790

Human Performance Effectiveness and Simulation

Approved for public release; distribution unlimited.

ARI Research Reports and Technical Reports are intended for sponsors of R&D tasks and for other research and military agencies. Any findings ready for implementation at the time of publication are presented in the last part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

---

## FOREWORD

The Simulation Systems Technical Area of the Army Research Institute for the Behavioral and Social Sciences (ARI) performs research and development in areas that include training simulation with applicability to military training. Of special interest is research in the area of simulation fidelity requirements.

Before any training system may be developed and procured for use in the Army training community its specifications must be determined. These training device specifications, when compared to the actual equipment may be defined as simulator fidelity. It is necessary to determine the effects of level(s) of fidelity on training effectiveness if guidance is to be provided to support fidelity decisions.

This report presents a conceptual framework for identifying factors that may impact training simulator effectiveness. A research strategy is proposed for the empirical determination of necessary levels of training simulator fidelity. The results of this report have implications for PM TRADE and for researchers in the areas of training and training device development.

JOSEPH ZEIDNER
Technical Director

ACKNOWLEDGEMENTS

The authors are indebted to their colleagues Dr. Richard E. Vestewig and Dr. Lee A. Miller, Honeywell, and to Dr. John J. Kane, Science Applications Incorporated, for their contributions to this report and their thoughtful comments on an earlier version of it. Thanks are also due to the Army Research Institute's Dr. Angelo Mirabella, Mr. Thomas Houston, and Dr. Robert Evans, who provided enthusiastic support and constructive criticism throughout the preparation of this report.

SPECIFICATION OF TRAINING SIMULATOR FIDELITY: A RESEARCH PLAN

BRIEF

---

Requirement:

To develop a research plan that can guide the determination of the empirical relationship between level of maintenance training simulator fidelity and training effectiveness. To, furthermore, ensure that this plan results in data that can be used to provide guidance for fidelity specification in the training simulator development process.

Procedure:

Three activities were undertaken at the outset of the effort: (1) a review of the empirical literature on training simulator fidelity was conducted; (2) data collection and interviews at selected Army, Navy, and Air Force agencies were carried out; and (3) a contract sponsored workshop entitled "Research Issues in the Determination of Simulator Fidelity" was held. A conceptual framework to guide this and subsequent efforts was developed. This framework identifies and defines factors of fidelity, utilization, and training effectiveness measurement that impact device training effectiveness. A research strategy was created based upon considerations involving (1) operational definitions of fidelity, (2) economic sequencing of studies, (3) theoretical and empirical issues of maintenance task training and required skills, and (4) desired subject population characteristics. Preliminary studies consistent with this strategy are proposed to explore the effects of device fidelity on the transfer of training of perceptual-motor and cognitive maintenance tasks.

Findings:

It is entirely feasible to conduct a systematic, empirical investigation of the relationship between level of training simulator fidelity and training effectiveness. This research can be carried out in such a way as to provide guidance in the context of the training simulator development process.

Utilization of Findings:

This report can be used by researchers to develop specific research plans for empirical studies to determine necessary levels of training simulator fidelity and by the military trainer development community to establish a framework for guidance in making fidelity decisions.

CONTENTS


| Chapter | | Page |
|---|---|---|
| I | INTRODUCTION AND BACKGROUND | 1 |
| | Program Objectives | 1 |
| | Purpose and Organization of this Report | 1 |
| | Statement of the Problem | 2 |
| | Methods | 3 |
| | Literature Review | 3 |
| | Site Visits and Interviews | 5 |
| | Fidelity Research Issues Workshop | 5 |
| | Factors Impacting Device Training Effectiveness | 7 |
| | Fidelity | 8 |
| | Definition of Fidelity | 8 |
| | Fidelity and Transfer of Training | 14 |
| | Fidelity and Fidelity Specification | 17 |
| | Utilization | 23 |
| | Training Effectiveness Measurement | 24 |
| | Summary | 26 |
| II | A FRAMEWORK FOR FIDELITY RESEARCH IN MAINTENANCE TRAINING | 27 |
| | Fidelity and Research Strategy | 27 |
| | Operational Definitions of Fidelity | 27 |
| | Sequence of Studies | 29 |
| | Fidelity Research and Maintenance Tasks | 30 |
| | Procedural Maintenance Tasks | 32 |
| | Perceptual-Motor Maintenance Tasks | 33 |
| | Cognitive Maintenance Tasks | 34 |
| | Fidelity Research and Subject Populations | 36 |
| | Summary | 36 |

CONTENTS (concluded)


| Chapter | | Page |
|---|---|---|
| III | PILOT STUDY DESIGN CONSIDERATIONS AND PARADIGMS | 38 |
| | General Design Considerations | 38 |
| | Micro Tasks | 38 |
| | Manipulation of Fidelity | 39 |
| | Transfer of Training Paradigm | 40 |
| | Research Paradigms | 40 |
| | Perceptual-Motor Task Pilot Research | 40 |
| | Cognitive Task Pilot Research | 46 |
| | System Selection and Task Design | 50 |
| | Summary | 51 |
| | REFERENCES | 53 |
| A | APPENDIX A. DESCRIPTION AND ANALYSIS OF THE ARMY'S TRAINING SIMULATOR DEVELOPMENT PROCESS | 59 |
| B | APPENDIX B. WORKSHOP DETAILS | 73 |

LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 1 | Sample Design for a Study to Determine the "Best" Combination of Similarity Levels for Training a Given Task. | 28 |
| 2 | Typical Configuration of a Bicycle Wheel. | 42 |

# CHAPTER I

# INTRODUCTION AND BACKGROUND

## PROGRAM OBJECTIVES

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has initiated a program of research on guidelines for training device and simulation development. This program, known as SIMTRAIN, has three major technical objectives.

1. Evaluate competing methods and models available for use in developing and evaluating training devices and determine appropriate applications in the existing acquisition process.

2. Develop guidelines for relating physical and functional training device characteristics (i.e., fidelity) to training effectiveness, with a focus on maintenance training.

3. Evaluate the training effectiveness of two alternative versions of the Army Maintenance Training and Evaluation Simulation System (AMTESS).

## PURPOSE AND ORGANIZATION OF THIS REPORT

This report presents interim findings and a research plan to initiate empirical efforts to develop data in support of the second objective listed above--specification of training simulator fidelity. The content of this report is based on a review of the relevant literature, data collection, and interviews at selected Army, Navy, and Air Force agencies, and the results of a contract sponsored workshop entitled "Research Issues in the Determination of Simulator Fidelity."

Chapter I describes these data collection and analysis activities, discusses issues related to the definition of fidelity and fidelity specification, and examines additional factors that determine device training effectiveness. Chapter II presents the rationale for a framework within which a specific plan of research can be developed and conducted. Chapter III presents a recommended series of studies to be undertaken within that framework.

## STATEMENT OF THE PROBLEM

It is widely recognized that simulators and training devices offer a potentially cost-effective alternative to training on actual equipment. The Army has an increasing commitment to replace or supplement hands-on training with training simulators. It is therefore necessary, in order to realize the potential increases in cost-effectiveness through simulation, to establish a systematically and empirically derived data base relating training simulator configuration and characteristics to training effectiveness.

Despite the many years of R&D on flight simulator fidelity, as simulation is applied to an increasingly wider range of tasks (e.g., equipment maintenance), there remain questions of what level of fidelity is appropriate, when to incorporate instructional features, and how to gain user acceptance.

Fidelity determination is a central issue in the specification and design of training devices and simulators. In general, the issue deals with the degree of fidelity necessary in a trainer to lead to a given level of training effectiveness. Although the issue appears unitary, it actually contains many components that have implications for all steps in the training device design and acquisition process, and for the overall training system for which the device is designed.

Considerable progress has been made in addressing the issues surrounding fidelity; however, much of the work on fidelity has been aimed toward specific areas of interest such as: particular devices, particular interpretations of the term, and particular models designed for fidelity determination. The work has so far lacked an overall programatic effort to unify the concept of fidelity, that

organizes existing work and indicates direction for future study. In addition, the work has not been aimed toward its application by the service/user community, and has not developed principles and guidelines that address the particular problems surrounding the specification of fidelity for maintenance trainers, as opposed to flight or team-trainers.

The current effort seeks to rectify this situation through a program of research designed to examine the general relationship between device fidelity and training effectiveness for the maintenance task domain. This report describes a plan to begin a programatic empirical research effort designed to provide data upon which guidance for fidelity decison-making by the training simulator development community may be based. The methods employed to help achieve these objectives are described next.

## METHODS

Three major activities were performed during the initial six months of the contract. These activities included:

- Literature Review
- Site Visits and Interviews
- Fidelity Research Issues Workshop

These activities are described in the following paragraphs.

### Literature Review

The objective of this activity was to create and then analyze and synthesize a bibliographic data base including design guides, technical reports, and academic and technology journal acticles. The review emphasized empirical as opposed to theoretical issues of training device fidelity and its relationship to training effectiveness. Both computerized documentation services and abstract publication lists were consulted.

The documents and other information judged relevant to this effort were obtained and cataloged. A review of the documents was performed and annotated abstracts prepared.

Each entry into the data base contained the following information:

- Internal document citation number (CN)
- Document's author
- Document's title
- Source
- File hardcopy availability
- Abstract
- Keywords

During the literature review, documents were abstracted and entered into a computerized data base. This data base is set up on Honeywell's Multics computer system. During the present SIMTRAIN effort additional document entries will be added to the data base.

Entering abstracts into the Multics data base offers several advantages. The computer can perform word processing, automated storage and automated retrieval. The primary advantage of the data base comes from automated retrieval through interactive queries. Keyword searches can be conducted in a number of ways, according to authors, sources, keyword list, or even keyword phrases contained in each abstract.

The information accessed in the search is printed by Multics in a form dictated by the searcher (e.g., alphabetized by author's last name, by consecutive citation number, etc.). In addition, at any time, the complete listing of all information in the data base can be printed off line for a complete hard copy of the abstract data base.

The abstract data base is intended to be a repository for information relevant to the development of guidelines for training simulator fidelity specification. It will be documented separately and updated throughout the course of this research program.

Site Visits and Interviews

The literature review effort was supplemented with site visits to selected Army, Navy, and Air Force agencies. These included:

- PM TRADE,
- Army Training Support Center (ATSC),
- Naval Training Equipment Center (NTEC),
- Navy Training and Analysis Evaluation Group (NTAEG),
- Navy Personnel Research and Development Center (NPRDC), and
- Air Force Human Resources Laboratory (AFHRL).

Discussions covered fidelity determination, training effectiveness, cost analysis, front-end analysis, specification development, joint working group decision making, media analysis, training delivery systems, individual differences, advanced technologies, user requirements, user acceptance, concurrent prime/training system development, and test and evaluation.

The Army training device development process is summarized separately in Appendix A to this document.

Fidelity Research Issues Workshop

In order to develop current positions on the research issues surrounding the determination of simulator fidelity, an invitational workshop was organized and conducted. The workshop format was chosen to take advantage of the expertise of the participants and to insure that issues critical to Army training requirements were addressed. Invitations were extended to nearly

100 individuals representing DoD, industrial, and academic organizations. There were 67 participants from 16 DoD agencies and 16 industrial companies and universities. A list of attendees will be found in Appendix B. Volunteers for presentations were solicited. The workshop was divided into presentation sessions and working group sessions.

Presentation Sessions--An initial presentation session provided an opportunity for ARI and SIMTRAIN contract personnel to describe the associated R&D program, and for selected participants to raise issues and offer perspectives on the problems of simulator design and utilization. A closing presentation session permitted the topic discussion leaders (see below) to summarize the results of the working group sessions and selected individuals to make closing observations and final remarks. Transcripts of the initial and closing presentation sessions will be published in a separate report.

Working Group Sessions--The working group sessions were structured so that each group had an opportunity to discuss each of four topic areas:

1. Effectiveness - covered issues of measurement and methodology in training and cost effectiveness and transfer of training.
2. Fidelity - covered issues of measurement and methodology in determining the relationship between physical/functional training device characteristics and effectiveness; also addressed the generalizability of previous research on flight simulators to simulators for technical (e.g., maintenance) training.
3. Guidance - determined appropriate formats and contents of guidance for fidelity specification decision making; also assessed when, how, and why to incorporate emerging technologies into simulators and training devices.
4. Priorities/Support - dealt with issues of topics 1-3 in terms of their priorities; also addressed the format and contents of agreements needed between R&D organizations and the sponsor and user communities to promote acceptance of R&D studies.

Each topic area had a topic discussion leader and an archivist; each group had a group leader. The groups sequenced through the topics so that only one group discussed a particular topic at one time. The initial working group session was three hours in duration, with the three subsequent sessions being one and one-half hours. Initial working group assignments matched individual interests/expertise to a topic area as closely as possible. The topic leaders facilitated intra-group communication.

The format chosen for the working group sessions was successful in stimulating discussions on the four topic areas. A wide diversity of views was expressed as to the appropriateness or form of the specific questions assigned to each topic area. Indeed, a large portion of Session I for all topics was devoted to refining or restating the questions. Therefore, not all of the questions posed at the outset of the session were discussed or answered. In some cases better (i.e., more meaningful) questions were the result. The main points arising from discussions of each topic group are summarized in Appendix B.

Chapter I concludes with a discussion of the fidelity concept and factors capable of influencing simulator training effectiveness. This background is based on an integration of the results of the foregoing activities. It is against this theoretical/conceptual background that we will develop a research framework and plan in Chapters II and III.

## FACTORS IMPACTING DEVICE TRAINING EFFECTIVENESS

The initial objective of the contract effort is to empirically determine the relationship between device fidelity and training effectiveness. This research must be carried out within a conceptual framework that will permit the development of guidance related to the many factors impacting device training effectiveness. Before substantial progress may be made in providing such guidance, the numerous variables which interact with simulator fidelity to produce given levels of training effectiveness must be empirically investigated. However, the empirical investigation of these interactions is beyond the

scope of the research described in this plan. It is the intent of the research described in this plan to construct a framework which will investigate the general relationship between simulator fidelity and training effectiveness and which will also afford a means for the systematic accumulation of data on all interactive variables which affect this relationship.

The basic form of a candidate model includes three factors:

- Fidelity
- Utilization
- Training Effectiveness Measurement

Such a preliminary model provides a framework for: (a) more rigorous modeling efforts; (b) research efforts addressing the value of the included factors, or addition/deletion of factors; and (c) measurement and validity issues involving various aspects of training simulator effectiveness. These three factors are discussed next.

## FIDELITY

Fidelity is a term long associated with training devices and their design (c.f., Miller, 1954). Fidelity has the status of an explanatory construct in attempts to account for or predict the training value of a device.

We next briefly review definitional issues regarding fidelity and select a definition for present purposes. Various theoretical perspectives on the relationship between fidelity and transfer of training are subsequently discussed. Finally, variables affecting the specification of fidelity are described.

### Definition of Fidelity

There is a clear lack ofconsensus in the literature on how to define simulation fidelity; this was evident at the workshop as well.

In general, definitions of fidelity suffer from the following deficiencies:

- The definitions are ambiguously stated, so that investigators use either different terms to describe the same type of fidelity, or the same terms to describe different types of fidelity.
- The definitions are not user-oriented; they give little guidance to the user on how fidelity may be determined or measured (Hays, 1980).

A major step in the development of a scientific approach is the definition of the concept of interest. Typically, definitions are initially informal, and are refined as additional data are gathered. Often, a definition of a concept such as fidelity must best be viewed as a "working" definition--admittedly lacking desired precision, but which can serve as an initial ground for generating hypotheses to refine the concept.

A working definition of fidelity must contain at least the following three components:

- Fidelity must be defined in terms of a domain of interest (X).
- Fidelity must be defined relative to something else (Y).
- Fidelity must be defined so as to be measurable.

A definition of fidelity must therefore be of the form "fidelity of X relative to Y as measured (or indicated) by Z procedure." If these components are present, then a working definition for fidelity can be formed.

The concept of fidelity, in itself, is not the major issue. Rather, it is the degree and type of fidelity of training devices relative to operational equipment that is required to lead to a given training outcome. Fidelity itself may be a superficial concept; as Matheny (1978) points out, a dictionary definition of fidelity means merely "duplication." The more significant issue is the departures from fidelity that can be undertaken in a simulator which will lead to a particular level of performance. Although strict

duplication can occur in only one way, departures from strict duplication can occur along many dimensions. Moreover, departures from absolute fidelity on one dimension do not necessarily covary with departures on other dimensions. Most definitions of fidelity that have been generated in the training community have been stated in such a way so as to emphasize either those dimensions which must be maintained tc leave the outcome of the training unaffected, or especially in the case of multidimensional definitions, those dimensions which can depart from duplication relatively independently.

Unfortunately, these definitions have often been confusing and contradictory among themselves, and lack the precision necessary to offer the user practical guidance.

The uses of the term "fidelity" have been numerous. In the area of trainers, two early uses of the term tend to accompany many of the later refinements of the term. R. B. Miller (1954) introduced the term "engineering fidelity" to describe the degree to which a trainer duplicates the physical, functional, and environmental conditions of the prime system. This purely hardware definition of fidelity is not necessarily adequate to encompass the training demands on the simulator. Consequently, the term "psychological fidelity" was introduced (Gagne, 1954) to represent the trainee's perception of the "realism" of the simulator. A distinction here is that the trainee may perceive a system that departs significantly from duplication as, nevertheless, highly realistic. The dimensions, therefore, may be, to some degree, uncorrelated. The distinction between these two types of fidelity is still present in contemporary formulations. Freda (1979) virtually duplicates these concepts in his two-component definition of fidelity. Physical fidelity is the "engineering representation" of the operational equipment, whereas psychological fidelity is indicated by behavioral and informational processing demands on the trainee. Freda points out that these are to be assessed independently, although both should be assessed in each device.

Many definitions of fidelity incorporate aspects of both engineering and psychological fidelity in their formulation. The classic three-component definition of fidelity of Kinkade and Wheaton (1972) includes 1) equipment

fidelity, which is the degree to which the simulator duplicates the appearance and control-feel of the operational equipment; 2) environmental fidelity, which is the degree to which sensory stimulation (excluding control-feel) is duplicated; and 3) psychological fidelity, which is the degree to which the trainee perceives the simulator as duplicating operational equipment. Both equipment and environmental fidelity are defined in terms of psychological variables, such as "feel" and "stimulus response" conditions, reflecting the importance of the training process in determining these. (It should be noted here, as in much of the work on fidelity, that the definitions, although meant to convey general principles, are drawn mainly from flight simulation, and, consequently, may contain features inappropriate to maintenance simulation. In the above formualtion, excluding control-feel from environmental fidelity is relevant in the sense that the flight simulator's controls are separate from the through-windscreen visual display. For maintenance training simulators, this distinction may be unnecessary and confusing.)

Condon, Ames, Hennessy, Shriver, and Seeman (1979) and Fink and Shriver (1978) offer formulations that incorporate both engineering and psychological fidelity. They use the term "physical fidelity" to describe layout and "feel" correspondence, and "functional fidelity" to describe stimulus/response correspondence. These dimensions are much like the Kinkade and Wheaton (1972) dimersions, except that they appear to contain a more "psychological" aspect; that is, both physical and functional fidelity are described in terms of the trainee's perception of them.

Condon, et al., point out that these can vary independently from each other, especially when type of task, sequences, or groups of task sequences differ. That is, depending upon which tasks are to be taught, physical and functional fidelity may be required to assume different values. The relevance of tasks is addressed also by Mirabella and his colleagues (Mirabella and Wheaton, 1975; Wheaton, Mirabella, and Farina, 1971); they introduce the term "task fidelity" to indicate task correspondence between the trainer and the operational equipment. This concept is very much like "behavioral fidelity"

mentioned by Condon, et al. The term "behavioral fidelity" may be preferred in some instances to an overall "psychological fidelity," since it is more descriptive of the highly behavioral, ISD-based approach to training that is presently demanded in today's military.

These conceptions of fidelity, in general, provide little guidance as to how they should be assessed in practical situations. Device design guide developers (Hirshfeld and Kochevar, 1979; Miller, McAleese, Erickson, Klein, and Boft, 1977; Marcus, Patterson, Bennett, and Gershan, 1980), modellers of predictive training effectiveness for device design (Swezey and Evans, 1980; Evans and Swezey, 1980), and manufacturers of training devices have chosen, in some cases, to avoid the term "fidelity" in favor of terms that are more closely related to the actual design process. Thus, physical similarity, functional similarity (Hirshfeld and Kochevar, 1979), and degree of correspondence of cues, responses, and actions (Miller, et al., 1977) describe fidelity-like concepts in terms more closely related to the judgments that must be made in device design. Physical and functional requirements (Swezey and Evans, 1979) emphasize the relationship of the design of the trainer to the training requirements which must be met. The use of fidelity-like terms for the practitioner, rather than more abstract definitions of fidelity, may aid in the practical work of trainer design and evaluation by not adding excess meaning of "fidelity" to the actual actions and conditions that must be met.

Hays (1980) performed a literature review on the concept of simulator fidelity, and, in general, reached the following conclusions:

- Many concepts have been used to describe fidelity in simulators. Although some investigators (e.g., Malec, 1980), take a very general approach to fidelity, most investigators view fidelity as consisting of various components.

- Investigators disagree on the important aspects of fidelity. The major disagreement seems to rest on the difference between the fidelity of device characteristics ("physical" or practical fidelity)

and the fidelity that the trainee experiences ("physical" vs. "behavioral" fidelity). Emphasis on each of these may lead to quite different device determinations.

- Investigators have developed terms for fidelity which overlap or are inconsistent with each other. The same type of fidelity is called by different terms, and alternatively different investigators use the same term in contradictory ways.

- Definitions are rarely operationalized to the degree that they can be unambiguously applied in specific situations.

- Definitions appear too "academic" in that they provide little guidance in themselves for the users of fidelity concepts, although guidebooks have been developed to provide "how to" information for the user.

- Fidelity determination must take into account issues such as stage of learning, trainee ability, general psychological principles of learning, the training content, and type of task.

Hays (1980) points out that previous definitions of fidelity have focused on both physical device characteristics and the trainees' perceptions, actions, and goals. He feels that these should be separate: "...fidelity should be restricted to descriptions of the configuration of the equipment and not be used when discussing behavior(s)" (p. 14). Fidelity concepts "become muddled as we attempt to use the same term to cover all aspects of the training situation" (p. 14). Behavioral and psychological implications of fidelity are important, but they should be described in such terms so as to not confuse them with device fidelity. The latter point is significant. It is necessary that fidelity be separated from the psychological implications of fidelity. This may be especially important for the device designer, who is necessarily more concerned with explicit device characteristics than with their implications.

Specifically, Hays (1980) proposed a definition of fidelity that has been subsequently refined as follows (personal communication):

> Training simulator fidelity is the degree of similarity between the training simulator and the equipment which is simulated. It is a two-dimensional measurement of this similarity in terms of:
>
> 1. The physical characteristics of the simulated equipment, and
>
> 2. The functional characteristics (i.e., the informational or stimulus and response options) of the simulated equipment.

Fidelity is thus a descriptor that summarizes the overall configuration and characteristics of a training simulator. All device features, those related to the actual equipment and those that are specific to the device (e.g., an instructional capability), impact the perceived fidelity of a training simulator. So, for example, a device specific feature that reduced fidelity in this sense could potentially enhance training effectiveness (Hays, 1980). It is this definition and sense of fidelity that we adopt for the purposes of the present program.

## Fidelity and Transfer of Training

Various basic theoretical approaches have been suggested to describe the relationships among fidelity, transfer of training and other related variables. Several of these recognized approaches are presented briefly below.

The Skaggs-Robinson Hypothesis (Robinson, 1927) was an early attempt to describe the relationship between transfer of training and similarity of training and operational equipment. This relationship is represented by a U-shaped curve where positive transfer decreases with decreasing similarity to a point, after which transfer becomes negative and then neutralizes as similarity continues to decrease. The implication for training device effectiveness, therefore, is that:

- High similarity results in high positive transfer;

- Moderate similarity results in negative transfer (due to confounding effects); and
- Low similarity results in essentially zero transfer.

The Osgood (1949) model of training transfer is perhaps the best known model which may be used to address variances in amount of transfer with gradients of similarity between the operational equipment and training devices. Osgood attempted to describe this relationship using a three-dimensional surface, relating input similarity (S) on one axis and output similarity (R) on the second, to amount and direction of transfer on the third.

According to the Osgood model, with identical stimuli, the effect of variation in required responses passes from maximum transfer at identical responses, through zero to negative transfer as antagonistic responses are reached. With identical responses, transfer drops to zero as stimulus similarity decreases. On the other hand, with antagonistic responses, transfer rises to zero from negative as stimulus similarity decreases.

Critics have found the Osgood model to be deficient in predicting negative transfer. Bugelski and Cadwallader (1956), for example, in testing this model obtained very similar results for positive transfer predictions, with contradictory results regarding negative transfer. The data obtained by Bugelski and Cadwallader appeared to conform much closer to the predictions based on the Skaggs-Robinson Hypothesis.

Miller (1954) attempted to describe the relationships between simulator fidelity and training effectiveness in terms of cost. The relationship posed by Miller is a hypothetical one which has never been substantiated by experimental or field data.

Miller hypothesized that as the degree of fidelity in the simulator (training device) increases, the cost of the training would increase as well. This relationship is represented by a exponentially increasing "cost" curve. An S-shaped "transfer" curve depicts the hypothetical relationship between fidelity level and transfer value. At low levels of fidelity, very little transfer value can be gained with incremental increases in fidelity. However,

at greater levels of fidelity, large transfer gains can be made for small increments in fidelity. The curve again levels off at high levels of fidelity, where further increments result in very small gains in transfer. Miller hypothesized a point of diminishing returns, where gains in transfer value are outweighed by higher costs.

Kinkade and Wheaton (1972) offered several general guidelines for choosing among alternative training device designs. The most significant guideline in their model highlights the non-absolute nature of simulator fidelity requirements. Kinkade and Wheaton (1972) point out that different degrees of equipment and environment fidelity (see Definitions of Fidelity) may be appropriate at different stages of training. They distinguish among three successive stages: procedures, familiarization, and skill. On this view, a single level of overall fidelity will yield differential amounts of transfer depending upon the stage of training.

In this light, it is important to point out that assessments of the physical and functional similarity of a device in comparison with the operational equipment provide only an indication of the realism of the training equipment and not necessarily its training effectiveness potential. There are a number of other variables that determine a device's training effectiveness.

According to the Advisory Group for Aerospace Research and Development (AGARD; 1980) for example,

> Greater value could be put on this measure (fidelity) if we could be certain that high fidelity is needed in every case in order that effective training and transfer of training can take place. However, such an assumption needs to be treated with caution. First, the amount of fidelity required will vary with the nature of the task to be trained, and so training effectiveness is likely to vary from subtask to subtask rather than be represented by one unitary value. Second, accepting that some fidelity is essential, it is still only part of the total training environment. The purpose for which the device is used, the form in which the training is given, the quality of the instruction, and the attitudes of the students and instructors toward synthetic training will all influence training effectiveness. Third, high fidelity, in assessing training effectiveness, is sometimes confused with the needs of the training environment itself. This

> state of affairs is no more clearly apparent than in the use of large simulators in commercial aviation....Training may be possible with far less sophisticated devices. Finally, while it would appear to be the case that high fidelity generates greater user acceptance, this fact does not of itself mean that a device is a more effective training facility (p. 35).

A number of these variables are or should be taken into account during the training simulator development process in which levels of fidelity are specified.

## Fidelity and Fidelity Specification

Fidelity has been previously defined as the physical and functional similarity of a training simulator with respect to the actual equipment. A number of variables affect the judged level of device fidelity and the specification of fidelity, including those related to the (1) task(s) to be trained; (2) the actual equipment characteristics; and (3) the requirements/characteristics of the training environment and personnel.

Task Variables and Fidelity Determination--Most authors agree (e.g., Hays, 1980; Wheaton, et al., 1976) that the question of training simulator fidelity must be considered in the context of particular tasks and equipments. The workshop participants concurred. Of particular interest is the influence of task variables on the determination of fidelity and on the configuration and characteristics of the resulting training device or simulator. A brief description is provided for eight task variables having a direct impact on device design and thus an indirect but important impact on device training effectiveness.

Task Domain--In the military environment where personnel interface with equipment, there are basically two task domains: operation and maintenance. Equipment operation may be further divided into two categories depending upon whether the equipment is moving under operator control (vehicles, aircraft, guns, and so on) or is stationary (display/control consoles, test equipment, etc.).

The job of a maintenance technician requires that he or she also be able to operate equipment (e.g., front-panels in the conduct of built-in tests). However, rarely does the equipment operation performed by the maintenance technician take place in the context of, for example, vehicle control (although a test ride does provide a counter instance). To the extent that different task domains or domain categories require different degrees of fidelity for effective training, the task domain will clearly impact fidelity specification. Furthermore, as noted earlier, fidelity may have a different definition in the context of a task associated with real-time, real world views and movements than in that of a task performed on stationary equipment.

Task Type--Perhaps the principal factor affecting the design of a training simulator is the match between the type of task being trained and the nature of the device. While one could enumerate a huge variety of task types (c.f., Fleishman, 1967), Card, Moran, and Newell (in press) have offered a rather cogent taxonomy for this purpose. They postulate the existence of three principal types of tasks: sensory/perceptual (those involving the input of information to the human processor), cognitive (those involving the internal processing of the observed input), and motor (those involving the human's processed output and their translation into observable actions). The design of a given training device must depend in large part on the extent to which it must allow for performance of a particular task type.

Task Difficulty--This variable (see Lenzycki and Finley, 1980, for example) relates to the performance difficulty of a task. It reflects both the task complexity and the adequacy of the work environment in which a task is to be performed. According to Lenzycki and Finley, there are four levels of task difficulty: unskilled (requires no training or experience to accomplish); easy to perform in the operational situation (equipment and work environment adequately designed; normal ambient conditions have no effect on performance); fairly hard to perform (some constraints in the operational envinonment); and hard to perform (the work environment, ambient conditions or the equipment design can produce major errors in task performance).

Task Frequency--This factor refers to the frequency with which a task is performed under operational conditions. Task frequency may have a paradoxical effect on the design of training simulators. While a frequently performed critical task may appear to warrant a high cost training device, the amount of practice afforded the individual in the operational situation may offset this requirement. However, the infrequently performed critical task may not be as frequently practiced and may require the more costly training device.

Task Criticality--Task criticality refers primarily to two characteristic components: (see Cristal, 1973, for example) delay tolerance and consequences of inadequate performance. Task delay tolerance addresses the amount of delay which can be tolerated between the time the need for task performance becomes apparent and the time actual performance must begin, whereas consequences of inadequate performance measures the impact of human error on the system.

Task Learning Difficulty--Another training simulator design variable involves the degree of learning difficulty associated with trainees' attaining a required level of proficiency for a particular skill or knowledge. Two task characteristics enter into such an analysis: the level of skill or knowledge proficiency necessary for task performance and the level of learning difficulty required to master the required skills or knowledges (see Wheaton, et al., 1976). As an example of the variables, four levels of task difficulty were identified by Swezey and Evans (1980): easy (trainee can accomplish this activity once informed that it exists; virtually no practice or study is required); modestly difficult (trainee can accomplish most of the activity subsequent to instruction with little practice or study, but some of the activity does require minimal practice or study to sustain competent performance at the desired level of proficiency); difficult (trainee can accomplish the activity following instruction, but only with consistent practice or study); and highly difficult (trainee requires extensive instruction, practice or study to accomplish the activity; requirement at least borders on expert performance standards).

Task Practice Requirements--According to Lenzycki and Finley (1980), the practice requirements of a task is the first criterion against which one must assess tasks to determine the need for a device as a medium for training. This variable addresses the extent to which initial practice and/or sustainment training are required to establish and maintain an acceptable proficiency level in task performance.

Task Required Skills, Abilities, and Knowledges--Dunnette (1976) has recently reviewed the literature in the areas of human skills, abilities, and knowledges. The establishment of what types of skills, knowledges, abilities, etc. are required by various tasks to be trained by a device is an integral component in addressing device design. For example, the ease with which tasks can be learned in a device and transferred to operational equipment varies with the nature of the task. Procedural skills will generally transfer readily but will be forgotten unless practiced regularly. Perceptual-motor skills transfer less completely because they are most susceptible to imperfections in the simulation of dynamic factors of environmental fidelity such as motion, visual, and kinesthetic cues and control forces. Nevertheless, while the level of transfer may be lower, rapid adaptation appears to take place in the operational environment. Devices for maintaining procedural skills are easier to design than are simulators to assist in the retention of perceptual-motor skill (AGARD-AR-159, 1980).

All of the above task variables must be considered in training simulator design. Typically such task variables are considered during the application of task analysis techniques. One reason for reduced training simulator effectiveness is the inadequacy of current techniques of task analysis as applied to training simulator development. An analysis of the tasks to be trained must result in a clear statement of what knowledges and skills have to be learned (or maintained) for job proficiency. There are several different task analytic frameworks or taxonomies in existence. Each one tends to have been designed for a specific purpose, e.g., training, operational equipment design, simulation model development, etc., or from a particular theoretical perspective, e.g., behavioral or cognitive (internal or mental mechanisms). As a recent experiment demonstrated, different taxonomies yield

significantly different results in simulations based on task-oriented network models of the same system (Berry, 1980). Clearly, then, there is the strong possibility that training simulators based on task analyses utilizing different taxonomies will yield different degrees of effectiveness. Shortcomings of a task analysis will promulgate through the development process and ultimately negatively impact force readiness.

Improved, tailored task analytic techniques are needed for the training simulator development process.

Actual Equipment Characteristics and Fidelity Determination--Representation of physical and functional aspects of actual equipment in a training simulator requires detailed knowledge of that equipment. That detailed knowledge comes from a system analysis which is usually carried out in parallel with a task analysis.

The functional aspects of equipment are typically more difficult to ascertain than the physical; this is especially evident in simulation of systems for maintenance training. For example, if high functional similarity is determined to be necessary, this implies that a working simulation (i.e., software) model of the actual equipment must be developed. Practical experience has shown the analysis and development of such a model to be a costly and time-consuming exercise.

The costs of achieving high physical or functional similarity can be extremely high. To achieve cost effective training we must be able to specify the minimum required level of fidelity to achieve a particular training purpose.

Training Environment/Personnel Variables and Fidelity Determination--There are a large number of variables associated with this factor. A minimal listing includes:

- Existing or projected training program constraints
- Device purpose
- Instructional principles

- Student population
- Instructor population
- Safety

Consideration of each of these variables will affect the eventual level of fidelity that a training device possesses. Instructional features and student population are briefly covered as representative examples.

Instructional Principles--Consideration of instructional principles in fidelity determination can result in the inclusion of instructional features in a training simulator. Instructional features are meant to enhance or facilitate the learning process, yet they lower the fidelity of the device because their incorporation reduces similarity of the device to the actual equipment. These features include, for example, augmented feedback or knowledge or results, performance measures and methods like adaptive training.

The relative contribution to training effectiveness of instructional features as compared to actual equipment features is unknown. Various authors and many of the workshop participants claimed that instructional features can account for a far larger portion of the variability in training effectiveness than actual equipment features. The approach adopted in the SIMTRAIN program is that this is an empirical question, i.e., one that should be answered through experiment. This research plan lays the foundation for this and other empirical efforts.

Student Population--In training simulator design it is important to consider individual differences in aptitude, ability, or skill-level. To the extent these variables are taken into account the fidelity of the resulting device may be different. For example, the needs of low aptitude students may result in the incorporation of more actual equipment features or functions, more extensive instructional features, or both. On the other hand, the needs of high aptitude individuals may result in fewer of these features being incorporated in a device. Once we have established baseline knowledge of the training effects of fidelity, we can begin to explore the potential interaction of fidelity with individual differences.

UTILIZATION

This is the least well understood and, according to some, the most potent factor in determining training simulator training effectiveness. A model of utilization factors should include user acceptance and motivation, and whether or not the device is used as it was intended.

The factors of user motivation and of user acceptance of training devices are of major interest in achieving effective training. These factors have been addressed by Burris, et al., (1971) and by Glaser and Resnick (1972), among others. According to AGARD-AR-159 (1980), device effectiveness is known to be highest when an instructor

> realizes and espouses the usefulness and relevance of the device even though he or she may be required to teach around faulty capabilities or features. Students tend to reflect instructor attitudes. Further, where recurrent or refresher training is being provided to experienced trainees, with or without instructor, an element of competition, or a comparative such as a probability of success in combat situations, tends to motivate trainees to better and faster learning. (p. 9)

A recent paper by Stoffer, Blaiwes, and Bricston (1980) presents a preliminary model of the acceptance process in research and development. A number of constraints are identified that can work against user acceptance of training R&D studies and training simulators. These are:

1. Deficiencies in user motivational conditions
2. Deficiencies in user role assignments
3. Deficiencies in official policy and structure
4. Inadequate defense R&D contracting methods
5. Inadequate integration of the user into the acquisition process through participative management
6. Other than rational user responses to R&D studies and to training simulators
7. Deficiencies in training simulator design

Training simulators have the potential to train. Students and instructors must use training simulators in such a way as to maximize that potential. The best design in the world will remain just that unless users accept a training simulator and fully incorporate it into their training program.

We must increase our understanding of how user acceptance, for example, impacts training simulator effectiveness. Increased understanding of the problem will lead to an ability to predict, control, and thus, improve it. As Stoffer, et al., (1980) state:

> Some may resist the idea that science can be used to analyze and actually influence something so apparently nebulous and subjective as "acceptance." Although the state of science is not well developed in the acceptance area, there are some theoretical and empirical bases for influencing levels of acceptance. An initial step in the scientific approach would be to document the extent of variation of acceptance found in various aspects of (naval) training. Factors that influence acceptance would subsequently be identified, described and prioritized for different applications. Then improved metrics for these factors would be generated. A conceptual framework consisting of these factors and their relationships to one another would be developed to understand the process of acceptance and to use as a basis for predicting acceptance levels in particular situations. Some of these factors have predictable, but uncontrollable consequences on acceptance levels; other factors are controllable by those with influence in the training command structure. The R&D community can become more proficient at managing the introduction of training innovations by applying those factors that can be controlled to influence acceptance. (p. 19)

The same approach could and should be followed to deal with other aspects of the utilization issue. Further analytic and empirical work based on these authors' model appears warranted.

## TRAINING EFFECTIVENESS MEASUREMENT

The previous sections describing fidelity and utilization covered independent variables that can influence training device effectiveness. What dependent variable(s) should be used to define training effectiveness? Answers to this question bear ultimately on whether the simulator is viewed as resulting in more effective training and improved on-the-job performance.

The choice of training effectiveness measures is not a simple one. Typically, the effect of the training device on eventual performance on the operational equipment is viewed as a two-stage process.

- Training effectiveness on the device. This is the degree to which training is accomplished on the training device, and is assessed by various measures of improvement on the device.

- Transfer of training to the operational equipment. This is the degree to which the training accomplished on the device generalizes to performance on the operational system.

In most cases training effectiveness and transfer of training issues are highly related. However, instances exist where the two should be treated separately. For example, the case where performance on the training device is excellent, but contributes little to performance on the operational equipment. Another possibility is where a device may be unpopular or lack face validity, but nevertheless leads to excellent performance on the operational equipment. Such relationships must be evaluated in the conduct of a program of research on simulation fidelity.

In general, three methods exist for the purpose of evaluating training simulator effectiveness. One involves a transfer experiment, where various items of training equipment are compared either to each other or to relevant operational equipment in terms of transfer-of-training. A second involves ratings, where various experts (for instance, pilots) are asked to rate training devices or simulators on their perceived utility for training (see Adams, 1979, for a discussion of these issues). A third method is the development of analytic models such as the TRAINVICE family of models (see Wheaton, et al., 1976; Swezey and Evans, 1980). In this approach, analytic techniques are employed to investigate the extent to which simulators or training devices adequately cover necessary training requirements, address relevant tasks, and/or provide for training on appropriate skills, knowledges, and abilities.

These methods are not without criticism. Adams (1979), for example, discusses shortcomings of both the transfer of training and ratings methods in the context of flight training simulators. Adams himself prefers an analytical approach based on whether or not the training device is designed according to reliable scientific laws of human learning. However, this is somewhat circular as the reliability of the scientific laws is dependent upon the demonstrated success of other devices designed according to the same principles.

It is our opinion that effectiveness data based on only one measurement technique is insufficient for research purposes. The converging evidence provided by using two or more of the methods will be invaluable to assessing the training effects of device fidelity.

## SUMMARY

The Army's goal of achieving cost-effective training through the use of simulators can be met through a sustained program of training device research. Because high fidelity training devices are costly, data are needed on the relationship between fidelity and training effectiveness. Many factors impact fidelity determination and training device effectiveness. A systematic, empirical investigation of these factors is necessary. In the following chapters we describe a framework within which this research can be carried out (Chapter II) and specific studies that should be undertaken (Chapter III).

# CHAPTER II

# A FRAMEWORK FOR FIDELITY RESEARCH IN MAINTENANCE TRAINING

The objective of the initial phase of our research is to conduct preliminary studies on the relationship between fidelity and training effectiveness for maintenance tasks. Any particular study should be carried out within a framework providing for a programmatic approach. The framework must include a definition of fidelity that can be operationalized and a strategy for exploring various levels of fidelity. The framework must be based upon a set of tasks representative of Army maintenance requirements and the results of previous empirical fidelity research with such tasks. It must provide a structure for the accumulation of future empirical data, such as the effects of potentially interactive variables. Finally, the framework must consider issues related to subject populations. In this chapter we develop the framework that will guide the specification of a program of research on training simulator fidelity. The research described here is designed to be the first step in establishing the required framework for future efforts.

## FIDELITY AND RESEARCH STRATEGY

Hays (1980) has described a framework for examining the effect of fidelity on training effectiveness. Following his distinction between physical and functional aspects of fidelity, he suggests treating these as separate factors in an experimental design. Figure 1 shows a sample design within this framework.

### Operational Definitions of Fidelity

Physical and functional similarity may be operationalized in terms of two scales (Wheaton, et al., 1971; Wheaton and Mirabella, 1972; and Mirabella and Wheaton, 1974). A panel layout index (PLI; Fowler, et al., 1968) can be used to determine physical similarity of a training simulator as compared to the actual equipment. Functional similarity can be assessed using the

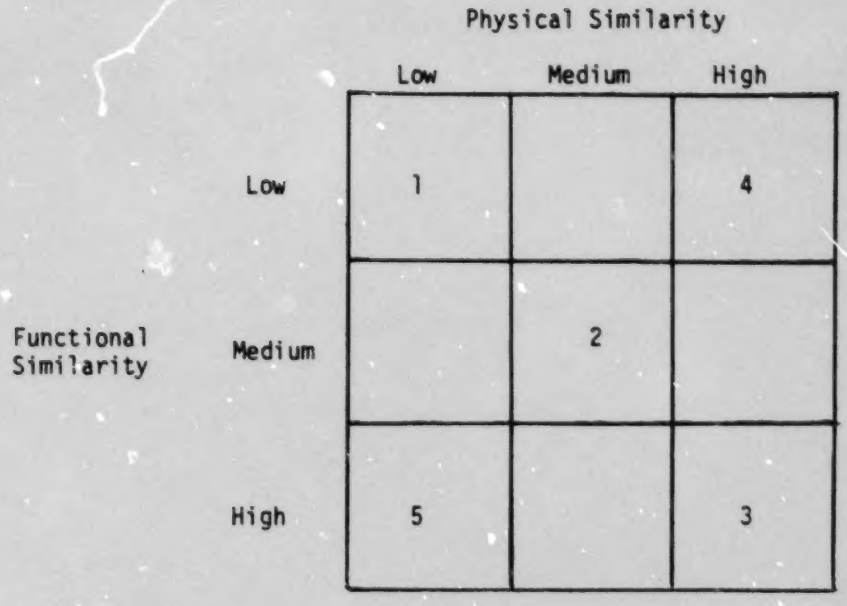



Figure 1. Sample Design for a Study to Determine the "Best" Combination of Similarity Levels for Training a Given Task. (After Hays, 1980) (see text for explanation of numbers in cells)

Display Evaluation Index (DEI) which measures the effectiveness of the display-control information flow (Siegel, et al., 1962).

A shortcoming of this approach is that the scales apply to equipment operation consoles, e.g., automatic test equipment, and would not be directly applicable to equipment lacking control panels and displays, e.g., a mechanical system. In these instances, observer judgment would provide a substitute.

The observer could be asked to rate physical and functional similarity on scales where each scale category was defined in terms of degree of similarity. Alternatively, a direct magnitude estimation procedure could be used. If enough alternatives were available, paired comparison estimates of overall fidelity could be obtained and the results analyzed using non-metric multidimensional scaling techniques (Shepard, 1964). In this manner the dimensionality of the space characterizing the alternative training simulators could be empirically determined. According to our current definition we would expect two dimensions.

Our approach will be to define fidelity in terms of a PLI and DEI or observer judgment of similarity, as appropriate.

Sequence of Studies

A potential problem with the design depicted in Figure 1 is that the two aspects of fidelity may not be entirely independent. For example, it might prove difficult or impossible to engineer a device that is low in physical similarity, while at the same time being high in functional similarity. In such instances we would have to extrapolate from conditions that could be achieved.

Additionally, achieving particular configurations of degrees of similarity could prove costly and the cost of providing nine alternatives might outweigh the benefit. Therefore, our strategy will be to take an incremental approach. Initially we will explore cells 1, 2, and 3 (see Figure 1) where high physical/high functional similarity will be represented by the actual equipment. Based upon significant effects along the diagonal we can then begin disentangling the effects of physical and functional similarity. This will be done by filling in the other cells beginning with 4 and 5 assuming they are technologically achievable.

## FIDELITY RESEARCH AND MAINTENANCE TASKS

In the previous section we developed a research strategy based upon considerations of how to operationalize simulator fidelity. The implementation of the strategy and the instantiation of levels of training simulator fidelity require the context of a specific task or tasks. We must have some criteria for determining which task(s) to study. In this section we consider criteria relating to the nature of equipment maintenance in the military, to generalizability of experimental results and to previous empirical research. In Chapter III criteria relating to the suitability of tasks for laboratory research are discussed.

The basic job of the equipment maintainer is to fault isolate and repair specific systems. Depending upon whether the system is electronic, mechanical, hydraulic, or hybrid in nature and whether the maintenance is carried out at the Organizational, Direct Support, General Support, or Depot level, the details of the job tasks change. These details minimally include:

- amount of troubleshooting skill needed;
- tools or test equipment utilized;
- degree of automatic testing and fault isolation provided; and
- actions required to remove/replace components or adjust, align, or calibrate them.

Given the differences that exist among maintenance tasks, how can we select tasks so as to (1) achieve generality and (2) maximize knowledge gained from previous research? Both of these objectives can be met if we accept a categorization of tasks based on skills required.

It is generally agreed in contemporary psychology that there are two broad categories of skill, namely, perceptual-motor and cognitive (Fitts and Posner, 1967; and Welford,1968). Both of these categories can be more finely divided. Perceptual-motor skill includes sensory, perceptual, and manipulative components. This category subsumes the Card, et al. (in press) sensory/perceptual and motor task types. Cognitive skill (Card, et al's. third task type) is exemplified by language comprehension, computation, decision-making, and problem-solving.

The performance of actual military maintenance tasks requires a combination of these skills. However, many complex tasks can be decomposed into their elemental components and upon so doing they can be seen to emphasize one type of skill as compared to another.

In the military equipment maintenance environment both kinds of skills are important. Perceptual-motor maintenance tasks, i.e., those requiring principally perceptual-motor skills, include such tasks as adjusting a carburetor, aligning radar synchronizers, and the like. Cognitive maintenance tasks include such tasks as interpreting oscilloscope waveforms, troubleshooting a system to the component level, and the like.

Other tasks require the integration of perceptual-motor and cognitive skills. In the military maintenance environment procedural tasks are the most important example of this integration.

There are thus three types of tasks that can be studied in the context of fidelity research. However, two issues remain: Which tasks do we already have good data on? What should our priorities be in selecting tasks to study? The following discussion provides answers to these questions.

Procedural Maintenance Tasks

Procedural tasks are those involving a preferred or proscribed sequence of events and actions. Procedures are performed either by following technical material, e.g., skill performance aids, or by recalling the steps from memory.

What do we know about the level of fidelity required to train tasks that are primarily procedural? Our review of the literature indicated that we know quite a lot. Numerous studies performed over the past 30 years have concluded that high training simulator fidelity is not necessary for the effective training of procedural tasks (c.f., Bernstein and Gonzalez, 1971; Grimsley, 1969a, 1969b; Cox et al., 1965; Prophet and Boyd, 1970; Mirabella and Wheaton, 1974; Crawford and Crawford, 1975; and Johnson, 1981). These studies have compared various low-fidelity training devices, e.g., photographs, charts, mock-ups and flat-panel simulators, with actual equipment and found them to be equally effective. It thus appears that physical similarity is more important than functional.

This general result has been obtained for both procedures following and recalling primarily in the context of vehicle and console operation, for example, start-up, check-out, and shut-down procedures. Such procedures typically involve switch setting, button pressing and the reading of discrete display information. To the extent that equipment troubleshooting involves procedures for equipment operation, e.g., use of (automatic) test equipment or running and interpreting built-in tests, the data indicate that training can be accomplished using low-cost devices with low functional similarity and moderate to high physical similarity.

Orlansky and String (personal communication) have noted that although this evidence cannot be denied, it has not had a major influence in the development of maintenance training simulators. Given every indication that this type of (front-panel) maintenance activity for electronics systems will continue to grow (c.f., Fink and Shriver, 1978; or Kenney, 1977) the maintenance training community should take a hard look at this deficiency.

As noted earlier, the performance of any particular procedure will require either or both perceptual-motor or cognitive skills for isolating malfunctions or repairing equipment. What do we know about the level of fidelity required to effectively train maintenance tasks when they involve behavior more complicated than switch-pressing and display reading.

## Perceptual-Motor Maintenance Tasks

In the maintenance environment, perceptual-motor skills are needed in adjustment, alignment, calibration, disassembly, or assembly. Intuitively, perceptual-motor tasks are those requiring "hands-on" experience for effective training. Therefore, the principle guiding selection of simulator configuration has been to provide the highest fidelity simulation achievable or the actual equipment in the event that technical or cost factors limit the achievable fidelity. This has been especially evident in the flight simulator area. Our literature review did not discover any empirical studies of deliberately reduced fidelity for training perceptual-motor tasks in the maintenance domain. This gap in the literature would seem to be an ideal one to fill. Chapter III of this report presents a paradigm for carrying out some necessary research.

## Cognitive Maintenance Tasks

The need for the equipment maintainer to think rather than follow or recall procedures can arise under a number of different circumstances.

- The observable symptoms of a malfunction must be recognized and interpreted in order to determine the applicable fault isolation procedure.
- Procedural steps often involve calculation or waveform interpretation.
- Inevitably procedures or automatic and built-in tests fail to always isolate the specific source of a system malfunction. This may be because procedures have been inadequately documented or because they isolate only to a (large) group of components. In the case of automatic tests it may be because they do not (or cannot) consider multiple or intermittent causes for a malfunction.

Under any of these circumstances the maintainer becomes a problem-solver or decision-maker. The observable symptoms of a malfunction are interpreted in light of what the technician knows about how the system is configured and how it works. In non-procedural troubleshooting the maintainer must decide what information, e.g., line voltage, absence of pressure, etc., is needed to eliminate or confirm suspect system components or subsystems. That is, he or she must in effect create a troubleshooting procedure.

According to a recent analysis by Bond and Towne (1979) the main difficulty in troubleshooting "is that the technician's cognitive map of essential physical relations (electronic, hydraulic, electro-mechanical, and so on) in a complex equipment is often incomplete, vague, or incorrect" (pp. 5-6). We are in agreement with this view, but would extend it to include the technician's usually faulty mental model[1] of how the system functions including the functional relationships among subsystems. These problems are compounded by the use of

[1] One can think of a mental model as a simulation residing in the technician's head. He can "run" the simulation to discover what might happen under certain circumstances.

suboptimal troubleshooting strategies even by relatively experienced trainees (Hunt and Rouse, 1981). For example, although the half-split troubleshooting technique[2] is taught in technical schools, it is not consistently used on the job.

We know surprisingly little about the relationship between device fidelity and training effectiveness for cognitive troubleshooting skills. Many trainers have been built over the last 15-20 years (Valverde, 1968 and Fink and Shriver, 1978 present lengthy reviews). These include flat-panel, computer graphics terminal and three-dimensional devices. Furthermore, computer simulation models have been developed to train the mental model aspects of electronic circuit troubleshooting (Brown, et al., 1975; and Brown and Burton, 1975), and general and system specific troubleshooting strategies (May, et al., 1976; Crooks, et al., 1977; Rouse, 1979 a, b; and Hunt and Rouse, 1981). The common thread tying all of these devices and simulations together is the relatively high degree of functional similarity present across varying levels of physical similarity.

In no instance has more than one type of device or simulation been compared to the same actual equipment for assessing training effectiveness. The Air Force Human Resources Laboratory, however, is currently conducting such an evaluation (Cicchinelli, et al., 1980). In general, the results of most "one on one" comparisons indicate that the training device is equally effective as the actual equipment for training (Fink and Shriver, 1978; Orlansky and String, personal communication).

These studies, however, are plagued by a number of weaknesses, some of which are enumerated by Cicchinelli (1980). Paramount among these is the inability to collect true transfer of training data because the actual equipment cannot be faulted. Most researchers (c.f., Miller and Rockway, 1975; Rigney, et al., 1978 a, b) are therefore limited to assessing device training effectiveness in

[2] In the half-split technique one attempts to successively reduce the size of the set of possible faulty subsystems or components in half through each measurement.

in terms of troubleshooting performance on the device itself.

We thus lack systematic and valid data on the device fidelity required for effective training of the various aspects of cognitive troubleshooting skills. Chapter III presents a preliminary research plan designed to provide the needed data.

## FIDELITY RESEARCH AND SUBJECT POPULATIONS

The target population (i.e., the one to which we seek to generalize our results) for the research effort is first-term, U.S. Army enlistees selected for maintenance job technical training. We will not be able to conduct our research using Army personnel. College sophomores, for example, may not represent a comparable group.

In order to insure full generalizability of the research results, we must select subjects as consistently as possible according to characteristics of the target population. These characteristics include demographic and individual difference variables. For example, ideally the subjects should have the same average education level and mechanical or electronic aptitude as Army maintenance trainees. In practice it will be impossible to match the subject and target populations on all variables or even every important variable. Rather we will have to prioritize the important variables and choose a subject population on this basis. The research described in this plan is designed to explore general relationships between simulator fidelity and training effectiveness. It will be the job of future research efforts to examine the effects of subject characteristics on this general relationship.

## SUMMARY

In this chapter we have developed a fidelity research framework through considerations of 1) operational definitions of fidelity, 2) economic sequencing of studies, 3) theoretical and empirical aspects of maintenance tasks and re-

quired skills, and 4) desired subject population characteristics. The research strategy and guidelines established will be instantiated in the research plan presented in Chapter III.

# CHAPTER III

## PILOT STUDY DESIGN CONSIDERATIONS AND PARADIGMS

In the previous chapter we developed a research framework through a consideration of strategy, tasks, and subject population characteristics. In this chapter we present potential research paradigms for studying the relationship between training simulator fidelity and transfer of training for perceptual-motor and cognitive tasks. Before describing the research paradigms, we turn to a discussion of general design considerations for the pilot studies.

### GENERAL DESIGN CONSIDERATIONS

There are three guidelines that will govern the design of our laboratory-based pilot studies.

- Study part- or micro-tasks.
- Determine appropriate manipulations of fidelity.
- Utilize transfer of training paradigm.

We next briefly discuss each of these guidelines before turning to a description of specific research paradigms.

#### Micro Tasks

The to-be learned tasks must meet certain criteria. They must be representative of the skills required in an actual maintenance task environment; task performance must be easily measured and the measurements must be valid, reliable, and sensitive; and the tasks must be learnable in a reasonable period of time. One way to meet these criteria is to study parts of tasks or micro-tasks rather than whole tasks. For example, cognitive tasks have structural (i.e., mental model) and process (i.e., troubleshooting strategy) components. Legitimate research questions can be raised and answered about each component.

Manipulation of Fidelity

As discussed earlier in this report, fidelity is a summary concept for characterizing the configuration and capabilities of a training simulator. In the proposed research we are primarily concerned with the effect of the (physical and functional) similarity between devices and the actual equipment on training effectiveness. Our aim is to hold other factors that may impact effectiveness, e.g., instructional method, constant, or, in the case of device instructional features, null in value. This is done as a first step. Future research efforts will build on the framework established here to examine the interactive effects of instructional method and other variables such as task type, task difficulty, device acceptance, and many others.

There are many choices to be made in the manipulation of fidelity. An overriding concern is that we define experimental conditions that differ significantly along defined dimensions of fidelity.

We have therefore defined a number of criteria or constraints that our manipulations of fidelity must meet. First, the manipulations must be consistent with the theoretical distinction between physical and functional similarity. Initially, both physical and functional similarity must bear the same general relation, i.e., high, medium, or low, to the actual equipment.

Second, the alternative configurations must be of interest to the Army for reasons of either cost or resolving decisons among competing technologies having a potential training application.

Third, the properties of the task that require learning via practice must be understood, the more thoroughly, the better. In this manner we can confidently manipulate the fidelity of those aspects of the actual equipment that are operative in skill acquisition.

Fourth, by the same token, the properties of the physical system embodied by the actual equipment must be understood so that functional characteristics can be specified and precisely manipulated.

Fifth, any alternative configurations chosen must be technologically and economically feasible.

Sixth, and finally, the manipulation of fidelity must be independently assessed through a consensus of observer judgment and ratings. We must have confidence that our low fidelity condition, for example, actually is low in fidelity.

Transfer of Training Paradigm

We are most interested in how various alternative device configurations differ with respect to conventional training, i.e., training with actual equipment. The key issue is to determine whether or how much training time can be saved by substituting a training device for the actual equipment. Criterion measurement must be on the actual equipment. The transfer of training paradigm is thus appropriate.

Each experiment will consist of an acquisition phase and a transfer phase. Acquisition will take place in the context of the actual equipment (control) or alternative device configurations. Transfer in terms of savings will be measured on the actual equipment. At later stages of the research the transfer conditions may be manipulated to reveal not only how much and what has been learned, but also how flexible or stable the learning is. For example, task difficulty could be operationalized as problem complexity or time stress. Time pressure would be an appropriate difficulty factor to manipulate because military skills ultimately must transfer to a war time context in which, e.g., time to repair equipment is critical.

RESEARCH PARADIGMS

The discussion in this section is organized around task type.

Perceptual-Motor Task Pilot Research

The goal of this line of research is to determine the effect of physical and functional training simulator similarity on transfer of training for perceptual-motor tasks. In the maintenance task environment, perceptual-motor skills are required in adjustment, alignment, assembly, disassembly,

and so on. Maintenance of mechanical systems, e.g., vehicles, accounts for a major Army training requirement. Training under conventional condition., using actual equipment, is costly and can be hazardous to equipment or personnel. Alternative training conditions are thus desirable.

For the purposes of this research we need a task that can be studied effectively in the laboratory, where the required performance is representative of Army maintenance tasks. We have selected the trueing of a bicycle wheel because it appears to meet these needs.

The wheel is a simple yet elegant physical system for translating power into motion. The geometry of a wheel and the relative rigidity of its surface and the surface on which it moves determine the efficiency of the power transformation.

Bicycle wheels should minimize weight. Support for the lightweight metal, semi-rigid rim is provided by (adjustable) spokes. In use, the geometry (i.e., shape) of this type of wheel changes, causing it to become imperfect. Because required power for motion increases with deviation from true, a great premium is placed on keeping these wheels true.

Trueing a wheel is not a simple matter. The task is complex enough to be frustrating to a novice, yet appears to be mastered in a reasonable amount of time--after trueing 5-10 wheels according to expert opinion.

The task consists of first detecting any misalignment (e.g., correctly attributing wobble to the wheel and not a loose axle), its location(s) and amount, and then manipulating the spoke nipples (see Figure 2 ) with a spoke wrench to correct it. Figure 2 shows the typical configuration of the spokes attached to a rim.

Misalignment is detected by spinning the wheel in the context of fixed reference points on either side of the rim (e.g., the brake pads if the wheel is on the bike). The principle involved in correcting the deviation is to loosen, via the spoke nipples, the spokes that go to the side of the hub that the rim

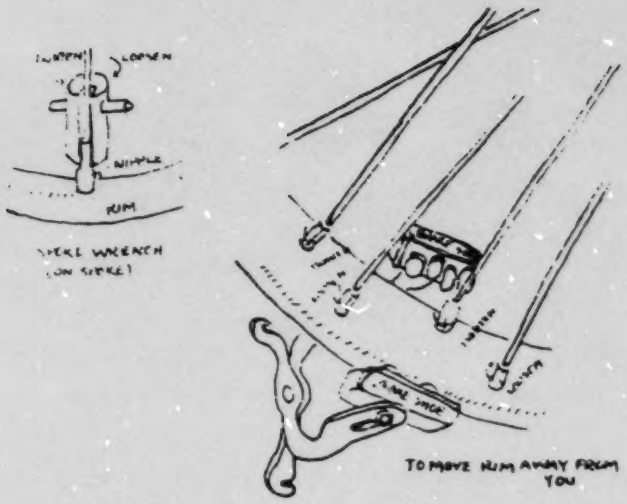



Figure 2. Typical Configuration of a Bicycle Wheel. (After Garvey, 1972). Illustrated are the spoke wrench and the principle involved in straightening the rim (see text).

pulls toward and tighten the spokes to the other side. This is a precision operation involving increasingly smaller adjustments of spokes farther away from the point of maximum deviation. The studies described next are aimed at discovering the effects of manipulating the fidelity of the training medium on the training transfer of a perceptual-motor skill.

Perceptual-Motor Pilot I--The purpose of this pilot study is to establish an expert baseline for the task. The temporal and performance measurement characteristics of the task will be determined. Several experienced individuals will participate as subjects. They will be required to true a wheel of known but variable deviation from true. A wheel stand will be instrumented for measurement using a travel dial indicator. The indicator measures degrees of displacement from a fixed reference.[1]

The results of this study will provide a characterization of the performance of a trained individual. Of particular interest are the length of time it takes to true the wheel under different amounts of deviation (difficulty), and whether the measurement technique described above needs to be supplemented by, e.g., observation of tool use.

Perceptual-Motor Pilot II--The purpose of this study is to establish the parameters of performance acquisition. The temporal and performance characteristics of task acquisition will be determined and a training method will be developed. Several naive individuals will participate as subjects. They will be required to learn how to true the wheel under conventional training conditions. Particular attention will be given to the method used for training and difficulty of the task under acquisition and transfer. The experimenter (E) will be experienced in the task and role play as an instructor during the acquisition phase. (It may prove difficult for E to perform both functions, in which case we will need a second individual to help run the study.) As the instructor, E will explain the principle of

---

[1] Note that spoke tension is also considered an important measure; it should be equally distributed. It is possible, but time consuming to measure spoke tension.

wheel trueing and demonstrate detection of wobble and use of the tool. The S will practice while E answers questions and provides feedback. This process will continue until a pre-established criterion (based upon the results of study I) is attained.

The results of this study will allow us to estimate the length of time needed for task acquisition. This data is critical for detailed planning of the remaining studies. Also, the data will indicate the shape of the learning curve for the task and provide an estimate of performance variability. These data will be critical to determining an appropriate level of difficulty of the task for training and transfer, and to estimating required sample size for the remaining studies. Finally, during the study, we will perfect and standardize the training methodology.

Perceptual-Motor Pilot III--The purpose of this study is to establish the general effect of fidelity on task training. Two simulated wheels will be evaluated as training (practice) media for task acquisition. One of the simulated wheels will be of medium fidelity and the other of low fidelity. Degradation of physical and functional similarity will be matched as closely as possible. Level of fidelity will be independently assessed through a magnitude comparison in which ratings of fidelity (in terms of physical and functional similarity) will be obtained from experienced individuals (subject matter experts).

The subjects in this study will be naive to the wheel trueing task. The training method will involve demonstration/observation and practice with feedback as described in study II. It is of utmost importance that the same method be used for training with both simulated wheels. Task difficulty levels during practice will be matched as closely as possible to one another and to that shown appropriate in study II. Task difficulty during transfer will be chosen so as to maximize the chance of finding differences between the alternative training devices.

The results of this study will establish whether we have an effective, i.e., significant, manipulation of fidelity. The data of this study can be analyzed in the context of data from study II, in order to provide an indication of the overall magnitude of the effect.

The nature of the pilot studies following this experiment will depend upon the results of Study III. Assuming positive results, Study IV will be as described next. However, negative results (lack of a fidelity effect) will cause us first to reexamine the fidelity manipulation, and the difficulty of the transfer task. Next we will change our selection of subjects who up until this point have been easily accessible (i.e., Honeywell employees) and unpaid for their services.

Perceptual-Motor Pilot IV--The purpose of this study is to replicate a portion of our previous results in the target population of interest. Pilot Study II will be repeated with refined training method and appropriate levels of difficulty. However, the subjects in this study will be naive individuals chosen from volunteer high school students. These volunterrs will be recruited from the population of students that will likely not continue on in their educatic..

The characteristics of this population can be established through demographic means. Liaison with a local high school will be established and within the limits of propriety and right to privacy, we will work with school officials to pinpoint the appropriate pool of students. If aptitude test scores are available and accessible, we will use them as a basis for identifying the population of interest. In this manner, a more representative, homogeneous group of subjects can be isolated. These subjects will be paid.

The reason for not using this population earlier is chiefly economic. A delay will also allow time for various administrative procedures and approvals necessary to gain access to the population of interest. For example, ARI and Honeywell must insure that the research methods and procedures are consistent with Government regulations for the use of human subjects.

Assuming positive results of this study, we will be in a position to collect data on the training effectiveness of alternative device configurations.

Perceptual-Motor Pilot V--This study will replicate Study III in which low and medium fidelity devices were used for training. High school students in a non-college track will serve as subjects. At this stage no finer categorization of subjects will be made.

Perceptual-Motor Pilot VI--The purpose of this study will be to determine if training effectiveness for this task is a differential function of physical or functional similarity of the simulated wheel(s) to the actual wheel. We will devise two additional devices, one of which will have high physical, but low functional similarity, while the other will be of low physical but high functional similarity. As in Study III we will obtain an independent assessment of the degree of similarity. In all other respects this study will be a replication of Studies III and V. At this stage in the research we should be able to use high school students as subjects. The results of this study will indicate the relative effects of both aspects of fidelity as determiners of training effectiveness for perceptual-motor tasks. Subsequent research can fill in additional levels/combinations of physical and functional similarity.

## Cognitive Task Pilot Research

The ultimate goal of this line of research is to determine the effects of varying levels of fidelity on the effectiveness of training troubleshooting skills. At the outset, however, the research effort must be directed at developing methodological and procedural tools for studying cognitive task acquisition and performance.

We have contended elsewhere in this report that cognitive task performance in equipment troubleshooting is based on (a) a cognitive model or representation of the system in question, and (b) a strategy for problem-solving. There are a large number of questions to be asked about both the structure (model) and the process (strategy). For example,

- How do we know if a model or strategy has been acquired? Is being used?
- How does the quality of the model or strategy vary with training conditions, e.g., fidelity?
- How do models or strategies vary with the nature of the system? The skill level of the user?

- Are some tasks better taught using models than others?
- How do we choose the appropriate model or strategy to teach?
- Are some strategies suitable to some models, but not others?

Prior to being able to carry out a series of studies as outlined in the previous section, we must develop a method for defining cognitive models and strategies and measuring the quality of decision-making in a cognitive task. Equally importantly, we must develop or select a system and design a trouble-shooting task suitable for laboratory study. In the near-term, we will proceed along two parallel paths directed at the methodological questions and the system/task questions.

Measurement Methodology Development--The purpose of this pilot study is to develop a method for assessing the quality of the models and strategies underlying performance in a complex, conceptual task. The method must provide data allowing inferences about internal structures and covert strategies and permitting discrimination among structures and strategies of different quality.

An analysis of troubleshooting cannot consist of a listing of procedures followed, since the skills of maintenance troubleshooting may not consist of following procedures. Skilled technicians appear to be using task and contextual information to make decisions and accomplish a variety of long-range and short-range goals. Accordingly, our methods will focus on the way skilled personnel use available information, and on the types of goals they are concerned with. If the methods evaluated are successful in accomplishing these objectives, we should be able to use them not only to make inferences about the quality of underlying models and strategies, but to provide guidance about the types of goals trainees must learn to be aware of and to use; the methods should also result in guidance about the type of decision options that must be represented by the training device in order for the trainee to learn to use goals to select between critical alternative pathways.

Therefore, the desired output for this line of research is a generalizable method that can ultimately be used to tell training device designers some specific requirements, in terms of goals represented and decision options presented. This guidance should constitute a partial resolution for questions of fidelity.

Rationale--The paradigm that will be assessed is a synthesis of work by Attneave (1954) and Shannon (1951) on information theory, with work by Klein and Klein (1981) studying goal networks of proficient CPR paramedics. Attneave (1954) discussed a paradigm in which a picture is constructed of three objects, each with a different color, and divided into a matrix of 50 rows and 80 columns, to produce 4,000 cells. The subject never sees this picture. The task is to guess the color of each cell. They predict one cell, are given feedback about the actual color, and move on to the next cell. Since only three colors are used, the probability of guessing correctly by chance is approximately 1/3 for each cell. Actually, subjects should only make 15-20 errors in the whole matrix, because they can use the redundancy of the picture (shapes, contours, symmetries). This paradigm was suggested as a way of exploring the information value of a visual figure, but it can also be used to probe a subject's ability to recognize and use redundancies. That is, if a complex domain is used, subjects with more expertise will be expected to make more accurate predictions about elements than subjects with little or no experience.

This method can only be used for tasks that can be analyzed into elements, but this would include tasks such as troubleshooting of equipment, where there is a sequence of tests and results. If the methods are fruitful, during Phase II of SIMTRAIN, we could analyze a troubleshooting problem into the tests and results sequentially obtained during the course of the diagnosis. The same problem would then be presented to a different technician, whose task would be predicting which test would be made next, followed by feedback on the results of that test, followed by a request to guess the next test performed.

This paradigm can tell us about the level of redundancy in the task, but that is not enough to understand training requirements. We need to understand the choices being considered. This is accomplished by asking the subject which are the likely alternatives at each choice point. By testing subjects at different levels of proficiency, we can identify changes in the types of alternatives they are considering to perform a task. Klein and Klein's (1981) CPR work has shown that more experienced personnel can be shown to utilize goal dimensions that less experienced personnel cannot

With regard to the question of fidelity, we hope to be able to ultimately use this paradigm to answer design questions about the number and types of alternatives that must be presented at choice points, for training personnel at any given level of proficiency.

The specific pilot study to be performed will evaluate the paradigm within the domain of chess-playing expertise. The game of chess was selected because it is a decision task that can be analyzed into elements. Subjects can be easily assigned to different skill levels, and there is good availability of motivated subjects. Finally, there are no equipment costs. All of these factors will allow us to perform the pilot study and analyze the results within available resources

Design--We will be contrasting three groups--experienced chess players (rated 1800-2100), average players (rated 1500-1600), and beginners (rated up to 1400). At least ten subjects will be run in each group. The materials will consist of three chess games actually played by players rated as experts (2000-2200). Each subject will attempt to predict the moves played for each of the games. For each game, the subject will be shown the chess board after move 10. Then, the subject will be asked which move was made by white on move 11. The subject will list the likely alternatives, as well as the actual guess. The subject will then be told if the guess was correct or, if not, which move was actually played. This process will continue for a sequence of 20 moves.

The data will be analyzed in several ways. First, we will compute accuracy of guesses by calculating the number correct out of the 20 moves studied. Second, we will compute the accuracy of subjects' ability to identify plausible options, by calculating the number of accurate alternatives listed, and dividing this by the total number of alternatives listed. This will enable us to see if accuracy of guessing increases with skill level, along with ability to recognize plausible options. By studying the patterns of novice and experienced chess players in generating options, we should be able to learn more about how this type of decision making task is performed.

Successful completion of this study and the analyses described will provide methodological tools useful for two purposes. First, using the prediction technique we should be able to derive implications about the simulation requirements for ensuring that an adequate range of alternatives is represented by a device.

Second, it should be possible to use these methods to evaluate the effectiveness of training devices and programs. Personnel at higher skill levels should show higher predictive accuracy. Once validated for chess we can apply the methods to cognitive troubleshooting tasks. The technique may be of particular value in tasks where there is no "right" answer; a measure of predictive accuracy will be a useful means of evaluating performance in non-procedural troubleshooting.

System Selection and Task Design

The purpose of this line of research is to assess the properties of different systems and tasks in light of the requirements of laboratory research. These requirements and some consideration of options for meeting them are briefly summarized below.

The system must be related in an obvious or, if possible, exact way to an actual Army system. Troubleshooting must be consistent with the maintainence philosophy of such a system. In short, both the system and task must have face validity and minimize artificialty.

The system must provide for a cognitively challenging task. Therefore, it must be relatively complex, consisting of subsystems built up of components, as opposed to a simple system consiting only of components. A complex system will provide the required task ambiguity and also the capability to vary difficulty -- troubleshooting can be accomplished at either or both the system and subsystem levels.

The system must lend itself readily to simulation. A system with digital signals would require less effort in this regard than one with analogue signals. However, an analogue system is more conceptually complex and may provide better discrimination of fidelity effects.

Learning the task cannot require lengthy participation on the part of the subject. Therefore, the troubleshooting task should be trainable with either minimal pre-training or pre-training associated with common courses of vocational technical study (e.g., basic electronics) so that a subject pool would be readily available.

Finally, the system and task should be accessible for replication in other laboratories. Therefore, nothing about either should be especially unique or require esoteric equipment, programming, skills or knowledge.

There are two candidate systems under consideration at present. One is a radio receiver, the other a microcomputer system.

In the course of this research we will explore the suitability of both by having one or two individuals become expert in their repair. These individuals will use a variety of techniques including interviews with experts and hands-on experience. The decision making method described in the previous subsection might also be used with experts.

SUMMARY

In this chapter we developed specific study designs for researching the relationships between training simulator fidelity and transfer of training.

for perceptual motor and cognitive maintenance tasks. The proposed experimental methods were based upon the need to (1) study part- or micro-tasks; (2) determine appropriate manipulations of fidelity; and (3) utilize a transfer of training paradigm. The perceptual-motor fidelity research will utilize a bicycle wheel trueing task; a series of six studies were described. The cognitive fidelity research requires the development of measurement method and selection of a specific experimental task; the necessary research and analysis was described.

## REFERENCES


Adams, J. A. On the evaluation of training devices. Human Factors, 1979, 21, 711-720.

Advisory Group for Aerospace Research and Development. Report of Working Group 10. AGARD-AR-159, December 1980.

Attneave, F. Some informational aspects of visual perception. Psychological Review, 1954, 61, 183-193.

Bernstein, B. R. and Gonzalez, B. K. Learning, retention and transfer in military training. Orlando, FL: Naval Training Device Center. Technical Report NAVTRADEVCEN 69-C-0253-1, September 1971.

Berry, G. L. Task taxonomies and modeling for system performance prediction. In Proceedings of the Human Factors Society, 1980.

Bond, N. A. and Towne, D. M. Troubleshooting complex equipment in the military services: Research and prospects. Los Angeles, CA: University of Southern California/Behavioral Technology Laboratories, Technical Report 92. December 1979.

Brown, J. S. and Burton, R. R. Multiple representations of knowledge for tutorial reasoning. In Representation and Understanding, Bobrow and Collins (Eds.) New York: Academic Press, Inc. 1975.

Brown, J. S., Burton, R. R., and Bell, A. G. SOPHIE: A step toward creating a reactive learning environment. International Journal of Man-Machine Studies, 1975, 7, 675-696.

Bugelski, B. R. and Cadwallader, T. C. A reappraisal of the transfer and retroaction surface. Journal of Experimental Psychology,1956, 52, 360-366.

Burris, R. W., McIntyre, M., Shaw, R. E., and Mast, T. Coordinating resources for a systematic approach to instructional research and development: An illustration in art history. A symposium presented at the Annual Meeting of the American Psychological Association, Washington, D.C., 1971.

Card, S. K., Moran, T. P., and Newell, A. The human information processor. In Card, S. K. and Newell, A. The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum, in press.

Cicchinelli, L. F. Factors limiting the measurement of maintenance simulator training effectivneess. Paper presented at the annual meeting of the American Psychological Association, 1980.

Cicchinelli, L. F., Harmon, K. R., Keller, R. A., and Kottenstette, J. P. Relative cost and training effectivness of the 6883 three-dimensional simulator and actual equipment. Air Force Human Resources Laboratory Report No. AFHRL-TR-80-24, September 1980.

Condon, C. F. N., Ames, L. L. Hennessy, J. R., Shriver, E. L., and Seeman, R. E. Flight simulator maintenance training: Potential use of state-of-the-art simulation techniques. AFHRL Technical Report: 79-19, Brooks Air Force Base, TX, 1979.

Cox, J. A., Wood, R. O., Jr., Boren, L. M., and Thorne, H. W. Functional and appearance fidelity of training devices for fixed-procedure tasks. Alexandria, VA: Human Resources Research Office, Technical Report 65-4, June 1965.

Crawford, A. M. and Crawford, K. S. Simulations of operational equipment with a computer-based instructional system: A low cost training technology. Human Factors, 1978, 20, 215-224.

Cristal, R. Comprehensive occupational data analysis profile (CODAP), USAFHRL Technical Report, 1973.

Crooks, W. H., Kuppin, M. A., and Freedy, A. Applications of adaptive decision aiding systems to computerized training system (Res. Pub. PATR-1028-77-1). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, January 1977.

Dunnette, M. D. Aptitudes, abilities and skills. In Dunnette, M. D. (Ed.), Handbook of Industrial and Organizational Psychology. Chicago, IL: Rand McNally, 1976.

Evans, R. A. and Swezey, R. W. Development of a user's guidebook for TRAINVICE II: Final report. Alexandria, VA: U.S. Army Research Institute, 1980.

Fink, C. D. and Shriver, E. L. Simulators for maintenance training: Some issues, problems and areas for future research (Res. Pub. AFHRL-TR-78-27). Lowry Air Force Base, CO: Air Force Human Resources Laboratory, July 1978.

Fitts, P. M. and Posner, M. I. Human Performance, San Francisco, CA: Brook Cole, 1967.

Fleishman, E. A. Performance assessment based on an empirically derived task taxonomy. Human Factors, 1967, 9, 349-366.

Fowler, R. L., Williams, W. E., Fowler, M. G., and Young, D. D. An Investigation of the relationship between operator performance and operator panel layout for continuous tasks. Report No. AMRL-TR-68-170. Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1968.

Freda, J. S. Fidelity in training simulation. Unpublished paper: Army Research Institute, 1979.

Gagne, R. M. Training devices and simulators: Some research issues. Technical Report: AFPTRC-TR-54-16, May 1954. Air Force Personnel and Training Research Center, ARDC, Lackland Air Force Base, TX.

Garvy, H. How to Fix Your Bicycle. San Francisco, CA: San Francisco Community Press, revised edition, 1972.

Glaser, R. and Resnick, L. G. Instructional psychology. Annual Review of Psychology, 1972, 23, 181-276.

Grimsley, D. L. Acquisition, retention and retraining: Effects of high and low fidelity in training devices. Alexandria, VA: Human Resources Research Office. Technical Report 69-1. February 1969(a).

Grimsley, D. L. Acquisition, retention and retraining: Group studies on using low fidelity training devices. Alexandria, VA: Human Resources Research Office. Technical Report 69-4, March 1969(b).

Hays, R. T. Simulator fidelity: A concept paper. Alexandria, VA: U.S. Army Research Institute, Technical Report 490, 1980.

Hirshfeld, S. F. and Kochevar, J. W. Training device requirements documents guide: A procedure handbook for Directorate of Training Developments Project Officers for Devices. Army Training and Doctrine Command, Fort Eustis, VA, 1979.

Hunt, R. M. and Rouse, W. B. Problem-solving skills of maintenance trainees in diagnosing faults in simulated power plants. Human Factors, 1981, 23, 317-328.

Johnson, S. L. Effect of training device on retention and transfer of a procedural task. Human Factors, 1981, 23, 257-272.

Kenney, J. W. Support systems for advanced military electronics. Autotestcon 77, 1977, 64-71.

Kinkade, R. G. and Wheaton, G. R. Training device design. In Van Cott, H. P. and Kinkade, R. G. (Eds.) Human Engineering Guide to Equipment Design (Revised Edition). Washington, D.C.: U.S. Government Printing Office, 1972.

Klein, H. A. and Klein, G. A. Perceptual/Cognitive analysis of proficient cardio-pulmonary resuscitation (CPR) performance. Paper presented at Midwestern Psychological Society Meeting, Detroit, MI, May 1981.

Lenzycki, H. P. and Finley, D. L. How to determine training device requirements and characteristics: A handbook for training developers. Research Product 80-25. U.S. Army Research Institute for the Behavioral and Social Sciences, May 1980.

Malec, V. M. Simulation in maintenance training. Unpublished paper: Navy Personnel Research and Development Center, San Diego, CA, 1980.

Marcus, G. H., Patterson, J. T., Bennett, C. D., and Gershan, B. S. Cost-effectiveness methodology for aircrew training devices: Model development and user's handbook. AFHRL Technical Report: 79-39, Brooks Air Force Base, TX, 1980.

Matheny, W. G. The concept of performance equivalence in training systems. In D. E. Erwin (Ed.) Psychological Fidelity in Simulated Work Environments; Proceedings of Symposium at the Annual Meeting of the American Psychological Association, Toronto, Canada, August 1978.

May, D. M., Crooks, W. H., and Freedy, A. Application of adaptive decision aiding systems to computer-assisted instruction: Experimental studies (Perceptronics Technical Report PATR-1023-76-1). Woodland Hills, CA: Perceptronics, Inc., March 1976.

Miller, G. G. and Rockway, M. R. Maintenance training in advanced instructional systems. In New Concepts in Maintenance Trainers and Performance Aids. NAVTRAEQUIPCEN IH-255, 1975.

Miller, L., McAleese, K., Erickson, J., Klein, G., and Boff, K. Training device design guide: The use of training requirements in simulation design. Air Force Human Resources Laboratory, Contract F33615-76-C-0050, June 1977.

Miller, R. B. Psychological considerations in the design of training equipment. Technical Report No. 54-563, Wright Air Development Center, Wright-Patterson Air Force Base, OH, December 1954.

Mirabella, A. and Wheaton, G. R. Effects of task index variation on transfer of training criteria. Orlando, FL: Naval Training Equipment Center, Technical Report NAVTRADEQUIPCEN 72-C-0126-1, January 1974.

Osgood, C. E. The similarity paradox in human learning: A resolution Psychological Review, 1949, 56, 132-143.

Prophet, W. W. and Boyd, H. A. Device task fidelity and transfer of training: Aircraft cockpit procedures training. Alexandria, VA: Human Resources Research Organization, Technical Report 70-10, July 1970.

Rigney, J. W., Towne, D. M., King, C. A., and Moran, P. J. Field evaluation of the generalized maintenance trainer-simulator: I. Fleet communications system. Technical Report No. 89. Los Angeles, CA: University of Southern California, Behavioral Technology Laboratories, October 1978(a).

Rigney, J. W., Towne, D. M., Moran, D. J., Mishler, R. A. Field evaluation of the generalized maintenance trainer-simulator: II. AN/SPA-66 radar repeater. Technical Report No. 90. Los Angeles, CA: University of Southern California, Behavioral Technology Laboratories, November 1978(b).

Robinson, E. S. The "similarity" factor in retroaction. Americal Journal of Psychology, 1927, 39, 297-312.

Rouse, W. G. Problem solving performance of maintenance trainees in a fault diagnosis task. Human Factors, 1979, 21, 195-203 (a).

Rouse, W. B. Problem solving performance of first semester maintenance trainees in two fault diagnosis tasks. Human Factors, 1979, 21, 610-618 (b).

Shannon, C. E. Prediction and entropy of printed English. Bell Systems Technical Journal, 1951, 30, 50-64.

Shepard, R. N. Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology.

Siegel, A. I., Mieha, W., and Federman, P. Information transfer in display control systems. VI. A manual for the use and application of the display evaluative index (Sixth Quarterly Report). Wayne, PA: Applied Psychological Services, 1962.

Stoffer, G. R., Blaiwes, A. S., and Bricston, C. A. User acceptance of R&D in Navy training: The problem major constraints, and an initial model of the acceptance process. Proceedings of the 1980 Industry Conference, Salt Lake City, Utah, 1980.

Swezey, R. W. and Evans, R. A. Guidebook for users of TRAINVICE II. McLean, VA: Science Applications, Inc., May 1980.

Valverde, H. H. Maintenance training media--an annotated bibliography. Technical Report AMRL-TR-67-154, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, OH, May 1968.

Welford, A. T. Fundamentals of Skill. London: Methuen and Co., Ltd., 1968.

Wheaton, G. R. and Mirabella, A. Effects of task index variations on training effectiveness criteria. Technical Report: NAVTRAEQUIPCEN 71-C-0059-1. Navy Training Equipment Center, Orlando, FL, October 1972.

Wheaton, G. R., Mirabella, A., and Farina, A. J. Trainee and instructor task quantification: Development of quantitative indices and a predictive methodology. Technical Report: NAVTRADEVCEN 69-C-0278-1. Naval Training Device Center: Orlando, FL, 1971.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., and Holding, D. H. Evlauation of the effectiveness of training devices: Literature review and preliminary model. Research Memorandum 76-6. U.S. Army Research Institute for the Behavioral and Social Sciences, 1976.

# APPENDIX A

# DESCRIPTION AND ANALYSIS OF THE ARMY'S TRAINING SIMULATOR DEVELOPMENT PROCESS

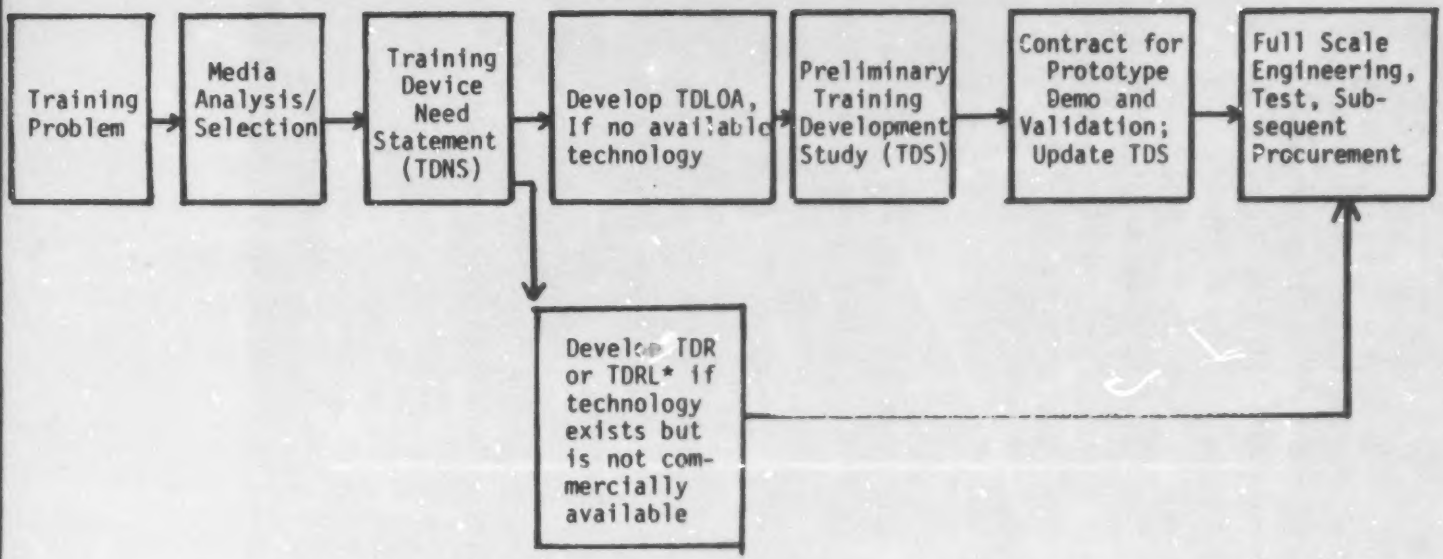
ARMY TRAINING SIMULATOR DEVELOPMENT

The following description of the Army's experimental training simulator development process is based upon the detailed artline found in TRADOC Circular 70-80-1 and the survey data collected during our site visits.

We describe the process herein so as to identify points that impact device configuration and capabilities. In this manner key leverage points for furnishing guidance can be isolated. These points are identified in a subsequent discussion of some of the strengths and weaknesses of the design process. A study approach for resolving some of the identified weaknesses is described.

There are two different sequences of action that may be pursued in the training device development. The one used depends upon whether the device will support a developing weapon system or whether it will support one or more fielded systems. It is usually in this latter instance, called nonsystem training device development, that PM TRADE becomes involved. However, at the request of the System Program Manager, PM TRADE can become involved in system training device development. Figure A-1 shows a highly simplified version of the nonsystem development process. The following description applies specifically to this example. Aspects of nonsystem and system training device development that impact fidelity specification are, in general, the same. The strengths and weaknesses of the processes in this report are thus highly similar.

The initiating condition for nonsystem training device development is the identification of a training problem. If a subsequent media analysis and selection indicates that a device is required to accomplish training


Training Problem
Media Analysis/ Selection
Training Device Need Statement (TDNS)
Develop TDLOA, If no available technology
Preliminary Training Development Study (TDS)
Contract for Prototype Demo and Validation; Update TDS
Full Scale Engineering, Test, Sub-sequent Procurement
Develop TDR or TDRL* if technology exists but is not com-mercially available

*Depending upon projected cost; see text for acronym meaning.

Figure A-1. Simplified Nonsystem Training Device Development (approval cycles not shown; commerical availability not shown; see Appendix C, Figure C-1 of TRADOC Cir 70-80-1 for more detail).

then a training device need statement (TDNS) is generated. The user (proponent school) is responsible for media analysis and selection. The TDNS is forwarded to the Army Training Support Center (ATSC) for action. The important parts of the TDNS from a device characteristics perspective are 1) a description of the nature of the training deficiency in terms of the training capabilities required, 2) an identification of the tasks to be trained, and 3) a description of the capabilities of the individuals who will be the subject of training with the device.

Upon submittal of the TDNS, ATSC coordinates with PM TRADE which determines the availability of existing technology (commercially available) to meet the training need. If an off-the-shelf solution or set of components does not exist, then ATSC tasks (and subsequently assists) the proponent with developing a training device letter of agreement (TDLOA) for the purpose of technology development. If technology is available then the proponent develops a training device requirements (TDR) document or a training device letter requirements (TDLR). (See below.)

There is much information contained in a TDLOA which will directly influence device configuration and characteristics. The format of the TDLOA is presented in Figure A-2. As can be ascertained from the TDLOA format, a considerable amount of training requirements analysis must be accomplished in order to provide the required information.

Upon submittal of a partial TDLOA, the proponent is authorized to establish a joint working group (JWG) for the purpose of finishing the TDLOA, and developing and coordinating additional requirements documents. The Principal Characteristics section is one highly important aspect of the requirements documents that is completed in the forum of the JWG.

1. Need.

a. A brief description of the training required to include the tasks to be supported, the current method used to train the tasks, the training system(s) to be replaced, and the time frame for which the new capability is needed.

b. Catalog of approved requirements documents (CARDS) reference number (to be assigned by DA ODCSOPS upon final approval of the document).

2. Operational concept.

a. A description of the role of the training device in the training concept and its relationship to other training systems and operational hardware.

b. The mission profile will be attached at Annex A. This annex should describe how the device will be used during a training day or days to accomplish training. It should describe a measure of how long the device will operate while being used to teach each task or series of tasks.

3. System description.

a. A statement indicating the principal characteristics expected to be included in the device, to include how the device will meet the training requirement, where it will be used, what the device looks like and those technological alternatives that have a reasonable chance of developmental success. Included, if applicable, must be any physical constraints or health hazards that will impact on design.

b. A discussion of other service, NATO/ABCA, or allied nation interest in the Army development/procurement. Include data on other service or allied developments with a view toward establishing potential for standardization/interoperability, or co-production. Include data on potential for procurement of allied nation items/systems.

4. Prospective operational effectiveness and cost. A realistic quantitative estimate of the training effectiveness that will be gained from the new training device when compared with the training system being replaced. This paragraph should include a subparagraph which identifies the estimated upper cost band where the training device fails to be cost effective.

Figure A-2. Training Device Letter of Agreement (TDLOA)

5. System development. This paragraph is divided into five paragraphs: operational, technical, logistical, training and manpower. Each subparagraph describes the device unique events which the training developer, materiel developer, logistician and administrator must undertake to insure the device is fielded as a safe and complete system. Include manpower constraints related to mission area or force level.

6. Schedule and milestones. A list of significant time phased events which will be conducted as a result of the TDLOA. As a minimum government acceptance of the prototype and DT/OT I, Validation IPR/ASARC II/DSARC II will be listed.

7. Funding. A broad estimate of the advanced development (AD), engineering development (ED), and unit flyaway costs. The AD and ED costs will be broken down by fiscal year and expressed in constant dollars provided by the materiel developer. This paragraph will also identify the number of prototypes which will be fabricated.

Annex A - Operational Mode Summary/Mission Profile. A list of ways the device will be used. For example, a device has the capability of instructing six hard-to-train tasks. The training developer must outline for what period of time each task is envisioned to be taught during the life of the device.

Annex B - Coordination annex. A list of organizations which provided comments with full rationale for not including LOA comments, if any.

Annex C - Rationale annex. Supports the various characteristics stated in the TDLOA.

Figure A-2. Training Device Letter of Agreement (TDLOA) (concluded)

The nonsystem training device development JWG is comprised of the following principal mandatory members:

- o Proponent TRADOC school (user/subject matter expert)
- o Army Training Support Center (ATSC-DST)
- o DARCOM/PM TRADE

As guidance to the proponent users, TRADOC Cir 70-80-1 lists 11 additional commands/activities that should be considered for participation in any JWG.

As noted above, the JWG completes the TDLOA during a series of meetings. Concurrent with this activity, a preliminary training development study (TDS) is developed. The TDS provides information on the probable cost- and training-effectiveness of the proposed device based upon task and media analyses, plus additional data resulting from, e.g., contractor studies or field use of similar devices.

Completion of the TDLOA and TDS initiates a series of staffing/approval actions which ultimately result in a demonstration and validation contract that is the responsibility of PM TRADE. The contract effort results in a prototype training device that is subject to developmental and operational testing (DT/OTI).

Upon completion of testing, or if technology is available, the proponent user and JWG must update the TDS and develop a TDR document or a TDLR, depending upon projected device cost. Subsequent to the completion of one of these documents, a series of staffing/approval actions takes place resulting in a full-scale development contract. If successful, as determined through developmental and operational testing (DT/OTII), the production and deployment of the device will follow.

The configuration and capabilities of the prototype and production training simulator are based upon a statement of work including an engineering specification prepared by PM TRADE.

In summary, there appear to be three categories of events in the device development process when decisions are made regarding device configuration/ characteristics (including fidelity). These are:

- o Media Analysis/Selection and TDNS
- o TDLOA/TDR/TDRL
- o Specification for RFP

STRENGTHS AND WEAKNESSES OF TRAINING DEVICE DEVELOPMENT

As can be observed from the description above and a review of TRADOC Circular 80-70-1, the training device development process is fairly systematic in terms of the sequence of decisions and required actions. However, although responsibility and authority are clearly defined in many instances, some ambiguities remain. The administrative mechanics of the process though detailed are not entirely consistent nor are they fully consistent with the Life Cycle System Management Model.

The individuals who participate in the process generally are motivated, dedicated, and competent within their areas of expertise. Specific problems do arise in the development process associated with each of the three categories of events listed above.

Media Analysis/Selection and TDNS

Media analysis is a part of the Instructional Systems Development (ISD) process based in part upon decisions made regarding learning objectives and classification of objectives into categories of learning. (A recent report

by Reiser and his associates at Florida State University (Reiser, et al., 1981) provides a current example of a media selection model.) If the ISD process has been used then this requirement presents no problem. However, rarely is the need for a (nonsystems) training device identified through a systematic analysis of the tasks (see previous section) to be trained. Therefore the information necessary to perform a media analysis is rarely available. Even if this information were available, the media selection procedure is highly subjective and requires judgments that imply a great deal of knowledge of cost and feasibility of alternatives, e.g., a training device or simulator. The analysis and selection of media requirements is a difficult task to do well. Even if a task analysis and learning objective categorization has been carried out, there can be problems because of the pitfalls inherent in task description. Some tasks are not readily described, especially those involving a significant amount of mental skill. Many task analyses will not reach the correct level of detail.

If the user has conscientiously and skillfully carried out media analysis and selection (and the implied ISD process) all of the information necessary for a TDNS will be available. Any weaknesses in the underlying analyses, however, will be promulgated into the device development process through the TDNS. Although there are sufficient safeguards and checks in the system to insure that a simulator or training device is necessary, at this point in the process it is difficult to determine if all tasks to be trained in fact require a device.

TDLOA/TDR/TDLR

These documents are grouped together because they describe device characteristics, whether the result of JWG discussions or prototype development and test. A major shortcoming in the preparation of these documents is that key members of the JWG lack specific guidelines for the types of decisions which they must make.

The JWG format was born out of the necessity to bring several different kinds of expertise to bear on the training device development process. However, the group format alone does not insure effective interaction among the users, training support personnel, and engineers. Each representative has his own biases and requirements and a language for describing what is wanted or needed. Two "decision rules," typically of the user, illustrate a costly bias:

- o In the absence of knowledge go to a higher fidelity, and
- o spend available money on fidelity until it is exhausted.

Fidelity decisions are driven by a desire first to use the actual equipment. Such a perspective also tends to ignore consideration of instructional features.

Group problem-solving is complicated enough, but when individuals representing different scientific, engineering and technical disciplines compose the group the task difficulty is compounded. The lack of a common conceptual framework and language for describing training simulator configuration and capabilities is the most significant impediment to intra-JWG communication. Improved communication among the JWG participants is necessary to insure that training requirements are efficiently translated into training simulator characteristics and ultimately engineering specifications.

Specification for RFP

Once the training characteristics and capabilities of a device have been specified, PM TRADE has the task of writing an engineering specification as part of a statement of work. This is a critical task in the training device development process and one that is difficult to do well. From the point of view of the engineer, he must preserve the integrity of the training requirements while at the same time detailing hardware, software, and interface capabilities. From the point of view of the device manufacturer, he must determine

how to implement the desired capabilities. Often times the same specification will be viewed as too detailed by one manufacturer and not detailed enough by another; the same can also be true for different parts of the same specification. From the point of view of the user, he must evaluate the engineering specification and determine if the resultant device will meet his training needs.

Standardization is coming to the process of preparing and documenting engineering specifications, but the pace of standardization must be quickened. The lack of a common, universally accepted (i.e., DoD-wide) framework for the language and format of training device engineering specifications is a shortcoming that must be resolved.

## A RECOMMENDED STUDY APPROACH

The goal of study in this topic area should be to elucidate the JWG decision-making process and to develop a framework for continued data collection. The objective of this line of research should be to determine the required format and content for guidance in the training simulator development process.

The method should consist of observation of JWGs concerned with the development of at least two, preferably three, different training simulators at different points in the development cycle. Data-gathering should be through note-taking and, when possible, follow-up interviews. The observer should not participate in the proceedings.

The analysis of the data should focus on the evaluation of "critical incidents." That is, those points of contention that in the judgment of the observer were not easily resolved. The incidents should be categorized and described. Particular emphasis should be placed upon the identification of the underlying cause of the incident. For example, was pertinent data unknown? unavailable? misinterpreted? Was sufficient detail available in task analyses? And so on.

The results of this study would form the basis for asking more specific questions about the JWG decision process and for designing field studies to answer them.

REFERENCES

Reiser, R.A., Gagne, R.M., Wayer, W.W., Larsen, J.Y., Hewlett, B.A., Noel, K.L., Winner, J.L. and Fagan, C. A learning based model for media selection Media selection flowchart and user's guide. Tallahassee, FL: Florida State University/Center for Educational Technology, 1981.

Training and Doctrine Command, Research, Development and Acquisition Training Device Development. Ft. Monroe; VA; TRADOC Air 70-80-1, October 1980.

# APPENDIX B

## WORKSHOP DETAILS

- o Agenda
- o Topic Area Questions
- o Discussion Summary
- o Attendees

WORKSHOP AGENDA

"RESEARCH ISSUES IN THE DETERMINATION OF
SIMULATOR FIDELITY"

23-24 July 1981

23 JULY

| | | |
|---|---|---|
| 0745 | Coffee and Registration | |
| 0830 | Call to Order | John Brock, Honeywell SRC |
| 0835 | ARI Welcome | Col Franklin Hart, Commander, ARI<br>Dr. Joseph Zeidner, Technical Director, ARI |
| 0840 | ARI Simulation Research Program | Dr. Angelo Mirabella, ARI |
| 0850 | SIMTRAIN | Dr. Stephen Hirshfeld, Honeywell |
| 0900 | Fidelity Determination--Task Objectives and Structure | Dr. David Baum, Honeywell SRC |
| 0915 | Workshop Objectives and Organization | Dr. Robert Hays, ARI<br>John Brock, Honeywell SRC |
| 0925 | Issues in Simulation | Dr. Robert Hennessy, National Research Council |
| 0940 | Constraints on Fidelity | Maj Jack Thorpe, DARPA |
| 0955 | A Fidelity Definition and Interactive Factors | Dr. Robert Hays, ARI |
| 1010 | Break | |
| 1030 | What We Know and Who Should Know It | Clarence Semple, Canyon Research |
| 1045 | Effectiveness Issues | Rohn Hritz, Applied Science Assoc. |
| 1100 | Definitions in Simulation | Worth Scanland, CNET |
| 1115 | Is Simulator Fidelity the Question? | William Montague, NPRDC |
| 1130 | Simulator Use in 4 Functional Areas | Dr. Leon Nawrocki, ARI |
| 1145 | Working Group Organization and Assignments | Dr. Robert Hays, ARI |
| 1200 | Lunch | |
| 1300 | Working Group Session I | |
| 1545 | Break | |
| 1640 | Working Group Session II | |
| 1730 | End of First Day | |

WORKSHOP AGENDA

"RESEARCH ISSUES IN THE DETERMINATION OF SIMULATOR FIDELITY"

23-24 July 1981

24 JULY

| | | |
|---|---|---|
| 0800 | Coffee and Rolls | |
| 0830 | Working Group Session III | |
| 1000 | Break | |
| 1015 | Working Group Session IV | |
| 1145 | Lunch | |
| 1245 | Informal Information Exchange/Topic Leaders Prepare Summations | |
| 1345 | Topic Leaders Present Summations | |
| 1500 | Break | |
| 1515 | Closing Observations | Dr. Jesse Orlansky, IDA<br>Dr. David Baum, Honeywell SRC |
| 1545 | Closing Remarks | Dr. Robert Hays, ARI<br>John Brock, Honeywell SRC |

I Effectiveness Topics

A. Training Effectiveness vs. Transfer Effectiveness

Is there a relationship between the scores of trainees on a training device/simulator (training effectivenss) and their scores on the actual equipment (transfer effectiveness)? If so: (1) How do we determine which scores or combination of scores to use in evaluating a training device? (2) How do we determine the nature of the relationship? (3) How might we apply test development methodology to develop reliable and valid evaluation tests?

B. Cost Effectiveness

(1) What is cost effectiveness?

(2) How do we measure it? Is there a cost effectiveness metric?

(3) How does cost effectiveness relate to training?

II Fidelity Topics

A. Measurement of Fidelity

(1) What is Fidelity?

(2) Is it measureable? If so, what kind of a metric? Why measure it?

(3) What is the relationship, if any, between fidelity and training?

B. Generalizability of Flight Simulation Data to other Areas of Simulation

Given that most data on the relationship between training simulator fidelity and effective training has come from the realm of flight simulation:

(1) Can we generalize flight simulation data to other areas of simulation training?

(2) What are the components/elements factors that influence generalizability?

(3) What are the conditions under which generalizability is possible?

(4) How do we use the flight simulation data that is generalizable?

III Guidance Topics

A. Development of a Format for Fidelity Decision Making Guidance

Given that the persons who must specify the characteristics to be incorporated into a training device need guidance in making fidelity decisions:

(1) On what issues/factors do they need this guidance?

(2) Is guidance available?

(3) In what format should this guidance be provided?

B. Impact of New Technologies on Fidelity Decisions

Given that new technologies are being developed which could impact on training strategies:

(1) How well do we use new technologies?

(2) How can we better incorporate new technologies into training systems?

(3) Is is possible/desirable for training devices to keep up with new technologies?

IV Priorities and Support Topics

A. Ranking Fidelity Research Issues in terms of Necessary Resources vs. Payoffs

Given that there are innumerable research issues relating to training simulator fidelity:

(1) How do we determine which issues to address and the order in which to address them?

(2) Do different groups rank issues differently? In which ways do they differ? Why?

B. Methods for Generating Long-Term Tri-Service Research Support and Communication

How do we gain the necessary support to undertake a long-term research effort to answer the questions raised in this workshop?

Support is necessary from the user level through DOD.

The format chosen for the working group sessions was successful in stimulating discussions on the four topic areas. A wide diversity of views was expressed as to the appropriateness or form of the specific questions assigned to each topic area. Indeed, a large portion of Session I for all topics was devoted to refining or restating the questions. Therefore, not all of the questions posed at the outset of the session were discussed or answered. In some cases better (i.e., more meaningful) questions were the result.

In what follows we enumerate the main points made within each topic area. These points generally represent a consensus of opinion, but where significant minority views were discussed they are noted as well.

EFFECTIVENESS

1. It proved difficult to establish general guidelines for defining measures of device training effectivneess. The particular measure chosen, whether normative (e.g., comparing a conventional training method with an experimental method) or criterion-referenced, depends upon the goals of the evaluation and of the particular training.

2. Are we concerned with terminal performance in a school setting or on-the-job? In a military context criterion-referenced performance, that is, ultimate performance on-the-job, is of paramount importance. A recognition of distinction between combat and peacetime performance of tasks is critical.

3. We must be able to measure effectiveness validly and reliably, before we can predict it. We do not always have the capability to measure performance, nor do we always know what to measure.

4. It is not possible to obtain combat performance measures in peacetime. We must test what we can test well and generalize or make assumptions based on sound psychological principles that other skills are trained well. Although there was general agreement that a transfer of training paradigm is the most appropriate for effectiveness measurement, there was a significant dissenting opinion that transfer, per se, does not allow one to isolate device effects.

5. Subject matter experts (SMEs) provide a valuable input on performance effectiveness, but we need better methods for obtaining information from SMEs.

FIDELITY

1. There was much discussion about whether we need the concept of fidelity at all. To make that determination the groups attempted to find a definition of fidelity that is task-independent. There seemed to be broad agreement regarding the definition of the physical aspect of fidelity. There was a consensus that (a) physical similarity was reasonably well understood; (b) it is important for some tasks but not for others; and (c) that non-physical similarity is often more important for training purposes. Opinions regarding the definition of non-physical (i.e., the other aspect of fidelity) were much more diverse and seemed to vary with the type of task under consideration. Major task types identified were procedural, control, and cognitive.

2. There was a consensus that fidelity is at best only one determinant of training effectiveness. Fidelity is a byproduct of the specification of the cues and responses required for task knowledge and skill acqusiition. This specification should result from an analysis of what is to be learned.

3. One can simulate an environment, equipment, tasks, the behavior of experts, and so on. Required fidelity depends in a complex way on the object of simulation and the training purpose.

4. It was pointed out that the instructors' needs are often ignored in a "fidelity" analysis. Instructor stations, or support functions, are either left out or poorly human factored.

5. A metric of fidelity should capture the categorical, hierarchical nature of training applications to particular tasks. Fidelity cannot be measured on a linear scale.

GUIDANCE

1. The idea of providing guidance on fidelity decisions, per se, was rejected in favor of an approach emphasizing a general class of design decisions. Devices must be designed with respect to training goals, including the nature of the tasks to be trained, personnel to be trained, level of proficiency needed, training context and methods, and so on. This general framework is not new and utilizing it in design avoids problems associated with focusing on only one aspect of the problem, namely, what level of fidelity to provide. A particularly interesting design factor that was suggested concerns the effectiveness evaluation criteria of a device; this was dubbed a "back-end" analysis. Attempts could be made to determine what is required for a device to be successful in the field and this information would constitute design criteria.

2. In the context of the device design decision process, fidelity is a byproduct of the training requirements analysis. It may not be useful to examine fidelity issues outside of this context.

3. It was universally agreed that we need better methods for communication between training psychologists and engineers. Present taxonomic approaches to describing and defining tasks/skills are limited and unwieldly.

4. The point at which training requirements are translated into engineering requirements is critical. There was much discussion about what format was best for presentation of training information to engineers. A good example of the dilemma is whether it should be in terms of the number of edges and corners for a visual display or in terms of the device being able to achieve 50% transfer. However, psychologists cannot tell engineers how many edges and corners are necessary without knowing the types of discriminations to be learned and a level of transfer of training cannot be specified without a complete account of performance measurement on the discrimination tasks.

5. Case studies were recommended as an approach for building a guidance data base. Documentation of lessons learned from device design decisions, utilization patterns and success/failure of specific features was encouraged.

6. Organizational changes were discussed as a key method to improving "corporate memory" which was judged to have improved somewhat with the experience of the last several years. Specific changes suggested included:

    - o providing decision-makers with direct experience on devices being used in the field;
    - o providing engineers with support to see devices they have designed in use;
    - o reducing turnover rates for training device developers by making it a career speciality;
    - o involving designers with engineering decisions, through procurement and several months of field observation before rotating back to design additional devices; and
    - o gearing rewards more to device training effectiveness rather than exclusively to budget and schedule factors.

7. There are many instances in which new technology, in the form of an added training capability, e.g., automated performance measures, freeze and replay, etc., has been misused or unused. Better communication is needed between designers and users, especially instructors who were not involved in the design process, as to the purpose and utilization of instructional support capabilities. Better methods for training of new users/instructors are needed.

8. More laboratory research is needed on the utilization and utility of new technologies before they are implemented in device design.

9. There is a lack of a focal point for the management of change. A facilitator, analogous to an agricultural extension agent, could be responsible for transmitting ideas and successes/failures between user community and engineers and training specialists.

PRIORITIES AND SUPPORT

The list of research issues identified is long and diverse. To complicate matters further, issues were presented at different levels of abstraction. Prioritization of the identified issues proved impossible during the group process, but the groups determined criteria that could be used to prioritize. Support was discussed in terms of two general categories: how to generate dollar investment for the research and how to facilitate information dissemination among services and disciplines.

The issues that seemed to generate the most discussion and general agreement were as follows:

1. Develop a "media selection model" for training simulator design.

2. Redirect the focus of R&D and the user community away from fidelity towards training effectiveness.

3. Identify and accommodate individual differences in aptitude, ability, or skill-level in training simulator design.

4. Determine how to coordinate the development of a training simulator's fidelity and instructional features.

5. Develop or utilize performance measures more closely related to real-world operational/combat criteria.

6. Develop techniques for training soldiers how to learn. Increase understanding of basic learning processes.

7. Survey existing simulators to determine if positive transfer is being achieved. To what degree? Can the relationship between device fidelity, features, and utilization and transfer be quantified? Over which factors does the device designer have control?

8. Develop a better model of the factors impacting device utilization. What are the correlates of user acceptance? Of increased student motivation? How do we cope with instructor turnover; training the instructor how to use the device?

9. To what extent do mental models underly task performance? What methods can be used to facilitate the acquisition of mental models?

10. Develop a technology for field-based experimental evaluations of training effectiveness. How do we intervene in ongoing curricula with new simulation ideas?

11. How do we promote/facilitate communication between the psychologist and the engineer? What are the job qualifications for a composite behavioral scientist/engineer?

The criteria against which issues could be prioritized included:

1. cost of the research
2. likelihood of effective payoff
3. feasibility of doing the research
4. sponsor priorities (i.e., policy)

5. technological gap closure
6. generalizability of the outcome
7. user acceptance
8. training effectiveness potential
9. availability of other necessary resources
10. sex appeal/PR value/visibility
11. operational requirements
12. low life-cycle cost of product of research

A peripheral issue involved the categorizatior of R&D and whether it is funded with 6.1, 6.2, or 6.3 monies: How can we better match research that is justified with the dollars available?

The support issues, as described, concerned resource allocation and dissemination of information.

1. Industry IR&D should be considered as a potential source of support despite the fact that companies will be reluctant to put money into endeavors that might benefit their competition.

2. More concrete evidence of the impact and payoff of training research is needed; especially in terms of benefits for the user community. The cost of training research could be put into the development costs of an operational system regardless of who paid for it. This could save money in the long run.

3. Application specialists (like Agriculture Field Service Representatives) for training could be established. They could facilitate during both the design and utilization phases of a training simulator's life cycle.

4. Remove limitations on the public distribution of Government studies to make literature more widely available.

5. Establish a centralized automated data base on training effectiveness. Such information as who is designing/building what devices, simulator descriptions, fidelity references, types of tasks trained, demonstrated transfer, etc. would be stored. The maintenance, update, and interpretation of such a data base would be difficult and great care would have to be taken in designing and implementing it.

ATTENDEES

ARI FIDELITY WORKSHOP

23-24 July 1981

Alluisi, Earl; AFHRL/CCN, AV 240-3605

Barber, Herbert; Ft Leavenworth, AV 552-5689

Baum, Dr. David R.; Honeywell SRC, 612-378-5048

Bessemer, David; ARI-Knox, AV 464-4932

Bickley, Will; ARI, AV 558-2873

Boldovicci, John; ARI, AV 284-8836

Boneau, C. Alan; George Mason Univ., 703-323-2203

Boycan, Gary; ARI, AV 284-9420

Brock, John F.; Honeywell SRC, 612-378-4403

Buckland, Maj George; AFHRL/OT, AV 474-6561

Campbell, Frank; Dir. of Training Development, Ft Rucker, AV 558-7111

Campbell, Dr. J. Olin; WICAT, 801-224-6400

Caro, Dr. Paul W.; Seville Research Corp., 904-434-5241

Chambers, Walter S.; NTEC, AV 791-5353

Chatelier, CDR Paul; OUSDR&E; 225-9777

Cicchinelli, Dr. Louis F.; Denver Research Inst/SSRE 303-753-2809

Clayton, Maj L. W.; USAF/RDIT, AV 858-4433

Clement, Warren; Systems Technology, Inc.; 415-961-4674

Connelly, Ed; Performance Measurement Assoc., Inc.

David, Elaine; Science Applications, Inc., 703-734-5872

Evans, Bob; ARI, AV 284-8836

Hagman, Dr. Joseph; ARI, AV 284-8695

Harris, Dr. Frank; ARI, AV 284-8827

Hays, Dr. Robert; ARI, AV 284-8836

Hennessy, Dr. Robert T.; National Research Council, 202-389-6064

Hirshfeld, Dr. Stephen; Honeywell SRC, 612-378-5377

Holman, Dr. Garvin; ARI, AV 284-9420

Houston, Thomas; ARI, AV 284-8836

Hritz, Rohn J.; Applied Science Assoc., 412-586-7771

Johnson, Dr. Richard; ARI, AV 284-8943

Kahn, Otto; ARI-Monterey, AV 929-8316

Kane, Dr. John; SAI, 703-627-4862
Kennedy, Dr. Bob; Canyon Research Group, 305-894-5090
Kimball, Gerry; Honeywell T&CSO, 213-331-0011 x 2818
Klein, Dr. Gary; Klein Assoc.; 513-767-2691
Lattimore, Andy; U.S. Army Training Support Ctr., AV 927-3564
Laughery, Ronald K ; Calspan, 716-632-7500
Laurence, Dr. Mike; ARI, AV 284-8836
Legum, Stan; Kinton, 703-820-1000
Magee, L. E.; DCIEM (Canada), 416-633-4240 x 306
Martin, Dr. Ed; ASD/ENETS, AV 785-4408
McLean, Craig; ASD/YWB, 785-2885
McMichael, Dr. Jim; Navy Personnel Research and Development Center, AV 933-7121
Mirabella, Dr. Angelo; ARI, AV 284-8827
Montague, Dr. William; Navy Personnel Research and Development Center, AV 933-7121
Muller, William; Honeywell SRC, 612-378-5086
Nawrocki, Dr. Leon H.; ARI Scientific Coordination Office, AV 680-3742
Olmstead, Dr. Joseph; HumRRO, 703-549-3611
Orlansky, Dr. Jesse; Institute for Defense Analyses, 703-558-1660
Pasnak, Dr. Robert; George Mason Univ., 703-250-6227
Peckham, Don; PM TRADE, AV 791-5881
Rizzo, Bill; NTEC, AV 791-5130
Root, Robert; ARI, AV 284-8721
Rose, Andrew; AIR, 617-275-0800
Scanland, Dr. F. Worth; CNET (Code N-5), AV 933-3591
Schendel, Dr. Joel; ARI, Ft Benning, AV 835-1360
Semple, C. A.; Canyon Research Group, Inc., 213-889-5072
Siecko, Nicholas A.; ECC, 215-687-2600
Smith, Deb; Honeywell, 612-378-5795
Smith, Ed; AFHRL/LRTR, AV 370-3391
Swezey, Dr. Robert; SAI, 703-827-4974
Thorpe, Maj Jack; Cybernetics Technology Div/DARPA, 202-694-1303
Tufano, Dr. Dan; ARI, AV 284-8836
Winter, David; AIR, 617-275-0800
Wood, LTC Piers M.; USATSC, AV 927-4681
Zacharias, Dr. Greg; Bolt, Beranek & Newman, Inc., 617-497-3397
Zeidner, Dr. Joseph; ARI, AV 284-8636

"END OF DOCUMENT"